

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



TRABAJO FIN DE MÁSTER

**Una herramienta para la extracción
dinámica y visualización de noticias,
comentarios y análisis de emociones en
periódicos online**

Máster Universitario en Ingeniería Informática

Autor: DE ANDRÉS NOGALES, Jorge

**Tutora: CARRO SALAS, Rosa María
Departamento de Ingeniería Informática**

JUNIO 2018

Resumen

En este Trabajo de Fin de Máster se propone la creación de una herramienta y una plataforma de soporte a la extracción y análisis del contenido emocional de los textos de noticias y comentarios de distintos periódicos online y a la visualización de toda la información obtenida. Además, brinda la posibilidad de añadir nuevos periódicos de manera dinámica utilizando la herramienta, pudiéndose incorporar toda la información correspondiente mientras se ejecuta la aplicación, sin necesidad de modificar la plataforma ni la herramienta en sí misma. Para ello, se ejecutarán unos scripts de manera externa que permiten obtener la información de los distintos periódicos, y a partir de ese momento la información queda integrada en la aplicación.

La herramienta permite descargar noticias de tres maneras distintas: mediante la introducción de la dirección URL de la noticia de un periódico, a través de los RSS del periódico y realizando una búsqueda con palabras clave entre las noticias del periódico. Esto expande las posibilidades de obtener la información que se necesita en cada caso, pudiendo obtener, respectivamente, una noticia en concreto, las noticias más recientes y las noticias relacionadas con un tema en particular.

Por último, se ha desarrollado una interfaz gráfica que permite realizar todas estas tareas y visualizar los resultados de manera sencilla. El usuario puede examinar todos los datos extraídos de cada periódico, incluyendo noticias y comentarios, y puede consultar los resultados del análisis del contenido emocional de los mismos. De este modo, es posible conocer las emociones que transmiten ciertas noticias, el estado emocional que provocan en la sociedad (reflejado en los comentarios de los lectores), la tendencia de unos periódicos u otros al transmitir las noticias, el impacto de las mismas, etc. Mediante la herramienta desarrollada se pueden analizar y comparar los textos de noticias y comentarios de distintas fuentes, estudiar los resultados del análisis de emociones de los mismos, analizar el impacto emocional de noticias similares en distintos periódicos, etc., de manera sencilla y rápida. La herramienta desarrollada se puede utilizar para analizar el *sentir* de la sociedad con respecto a temas de actualidad publicados en los periódicos online.

Palabras clave

Extracción dinámica de información, interfaz gráfica, Impacto emocional, análisis de emociones, análisis sentimental, periódico online, extracción de emociones.

Abstract

In this Master's Final Project is proposed the development of a tool and a support platform for extracting and analyzing the emotional content of the of news and comments text for different online newspaper and to the visualization of all the obtained information. It also brings the possibility of dynamically add newspapers using the tool, being able to add all the appropriate information while the application is running, without the need of modifying the platform or the tool itself. To do so, newspapers need some external scripts which will allow to obtain the information of the newspapers, and from this moment the information is stored in the application.

The tool allows downloading news in three different ways; by putting the news URL link of a newspaper, through the RSS of the newspaper and news keyword searching in the newspaper. This expands the possibilities of obtaining the needed information in each case, allowing to obtain a certain news, respectively, the breaking news and the news related to a topic.

Finally, a graphical interface has been developed that allows to perform all these tasks and visualize the results in a simple manner. The user can examine every extracted data from every newspaper, including news and comments, and can check the emotional analysis results of the news and comments texts. In that way, it is possible to know the emotions certain news transmit, the emotional state that provokes in society (reflected in the comments of the readers), the bias of some newspapers or when transmitting news, the impact of the the news, etc. Using the developed tool, it is possible to analyze and compare the texts of news and comments from different sources, study the results of the emotional analysis of the texts, analyze the emotional impact of similar news from different newspapers, etc., in a simple and quick way. The developed tool can be used to analyze the feeling that society has regarding to present topics published in online newspapers.

Keywords

Dynamic extraction of information, Graphic interface, Emotional impact, emotional analysis, sentiment analysis, online newspaper, emotion extraction.

Índice

| | |
|--|-----------|
| 1 INTRODUCCIÓN..... | 1 |
| 1.1 MOTIVACIÓN | 1 |
| 1.2 OBJETIVOS..... | 1 |
| 1.3 ESTRUCTURA DEL DOCUMENTO..... | 1 |
| 2 ESTADO DEL ARTE | 3 |
| 3 ANÁLISIS | 5 |
| 3.1 DESCARGA DE INFORMACIÓN DE LOS PERIÓDICOS ONLINE | 5 |
| 3.2 EXTRACCIÓN DE INFORMACIÓN DE MANERA GENÉRICA | 6 |
| 3.3 MÉTODOS DE OBTENCIÓN DE NOTICIAS | 6 |
| 3.4 FUNCIONALIDAD Y VISUALIZACIÓN DE INFORMACIÓN..... | 8 |
| 3.5 ANÁLISIS DE EMOCIONES | 8 |
| 3.6 REPRESENTACIÓN DE EMOCIONES..... | 9 |
| 3.7 ALMACENAMIENTO DE DATOS | 9 |
| 3.8 REQUISITOS | 10 |
| 3.8.1 Requisitos funcionales..... | 10 |
| 3.8.2 Requisitos no funcionales..... | 12 |
| 3.9 TECNOLOGÍAS A UTILIZAR..... | 13 |
| 3.9.1 Lenguaje y entorno de programación | 13 |
| 3.9.2 Descarga de información de páginas web | 13 |
| 3.9.3 Analizador de código HTML | 14 |
| 3.9.4 Interfaz gráfica..... | 15 |
| 3.9.5 Almacenamiento de datos | 15 |
| 3.9.6 Análisis de emociones | 15 |
| 3.9.7 Representación de emociones..... | 16 |
| 4 DISEÑO | 17 |
| 4.1 ARQUITECTURA GENERAL DEL SISTEMA | 17 |
| 4.2 MODELO - ESTRUCTURA DE DATOS..... | 19 |
| 4.2.1 Diagrama de clases del Modelo | 19 |
| 4.3 VISTA - INTERFAZ GRÁFICA | 21 |
| 4.3.1 Diagrama de clases de la Vista | 22 |
| 4.4 CONTROLADOR – LÓGICA DE APLICACIÓN | 27 |
| 4.4.1 Diagrama de clases del Controlador | 28 |
| 4.5 PERIÓDICOS | 30 |
| 4.5.1 Periódicos específicos..... | 30 |
| 4.5.2 Periódicos genéricos..... | 31 |
| 4.5.3 Scripts de descarga | 31 |
| 4.6 OBTENCIÓN DE NOTICIAS Y ANÁLISIS DE EMOCIONES | 33 |
| 4.6.1 Obtención de noticias..... | 33 |
| 4.6.2 Análisis de emociones | 34 |
| 4.7 INTERFAZ GRÁFICA DE USUARIO..... | 36 |
| 4.7.1 Ventana principal y menús..... | 36 |
| 4.7.2 Gestión de periódicos..... | 37 |
| 4.7.3 Añadir noticias | 39 |
| 4.7.4 Visualización de datos..... | 39 |
| 5 DESARROLLO..... | 41 |
| 5.1 PROYECTO DE JAVA | 41 |
| 5.2 CLASE PRINCIPAL..... | 41 |

| | |
|--|-----------|
| 5.3 MODELO | 41 |
| 5.4 VISTA | 42 |
| 5.4.1 <i>El panel central</i> | 42 |
| 5.4.2 <i>El panel visualización</i> | 42 |
| 5.5 CONTROLADOR..... | 43 |
| 6 EJEMPLOS | 45 |
| 6.1 DESCARGA DE NOTICIAS DESDE UNA BÚSQUEDA | 45 |
| 6.2 COMPARACIÓN DE NOTICIAS SIMILARES EN PERIÓDICOS DISTINTOS..... | 46 |
| 6.2.1 <i>Noticia en el periódico 20 Minutos</i> | 46 |
| 6.2.2 <i>Noticia en el periódico El Mundo</i> | 47 |
| 7 PRUEBAS Y RESULTADOS | 49 |
| 7.1 PRUEBAS FUNCIONALES..... | 49 |
| 7.1.1 <i>Pruebas de extracción de información</i> | 49 |
| 7.1.2 <i>Pruebas de interfaz</i> | 50 |
| 7.1.3 <i>Pruebas de gestión de periódicos</i> | 50 |
| 7.1.4 <i>Pruebas de almacenamiento y datos</i> | 52 |
| 7.1.5 <i>Pruebas de análisis de emociones</i> | 52 |
| 7.2 PRUEBAS NO FUNCIONALES..... | 52 |
| 7.2.1 <i>Pruebas de rendimiento y estabilidad</i> | 52 |
| 7.2.2 <i>Pruebas de usabilidad</i> | 52 |
| 7.2.3 <i>Pruebas de mantenibilidad</i> | 53 |
| 8 CONCLUSIONES..... | 55 |
| 8.1 CONCLUSIONES..... | 55 |
| 8.2 TRABAJO FUTURO | 55 |
| 8.3 CONSIDERACIONES FINALES..... | 57 |
| REFERENCIAS | 59 |
| A MANUAL DE USUARIO | 61 |
| A.1 INSTALACIÓN DE PROGRAMAS ASOCIADOS NECESARIOS | 61 |
| A.1.1 <i>PhantomJS y CasperJS</i> | 61 |
| A.2 UTILIZACIÓN DE LA HERRAMIENTA | 61 |
| A.2.1 <i>Sistema</i> | 61 |
| A.2.2 <i>Periódicos</i> | 62 |
| A.2.3 <i>Noticias</i> | 65 |
| A.2.4 <i>Datos</i> | 67 |

Índice de figuras

| | |
|--|----|
| ILUSTRACIÓN 1. ESQUEMA DE EXTRACCIÓN DE INFORMACIÓN DE MANERA GENÉRICA..... | 6 |
| ILUSTRACIÓN 2. URL RESULTANTE TRAS REALIZAR UNA BÚSQUEDA EN EL PERIÓDICO ONLINE 20 MINUTOS | 7 |
| ILUSTRACIÓN 3. ESQUEMA DE MODELO-VISTA-CONTROLADOR | 17 |
| ILUSTRACIÓN 4. ARQUITECTURA GENERAL DEL SISTEMA | 18 |
| ILUSTRACIÓN 5. DIAGRAMA DE CLASES DEL MODELO | 19 |
| ILUSTRACIÓN 6. DIAGRAMA DE CLASES DE LA VISTA | 22 |
| ILUSTRACIÓN 7. DIAGRAMA DE CLASES DEL CONTROLADOR | 28 |
| ILUSTRACIÓN 8. FORMATO JSON DE SALIDA PARA SCRIPT DE DESCARGA..... | 32 |
| ILUSTRACIÓN 9. FORMATO JSON DE SALIDA PARA SCRIPT DE RSS | 33 |
| ILUSTRACIÓN 10. FORMATO JSON PARA SCRIPT DE BÚSQUEDA | 33 |
| ILUSTRACIÓN 11. FORMATO JSON DE ENVÍO DE TEXTOS PARA ANÁLISIS DEL SENTIMIENTO..... | 35 |
| ILUSTRACIÓN 12. FORMATO JSON DE RESPUESTA DEL ANÁLISIS DEL SENTIMIENTO DE TEXTOS ... | 36 |
| ILUSTRACIÓN 13. VENTANA PRINCIPAL DE LA INTERFAZ GRÁFICA DE USUARIO..... | 37 |
| ILUSTRACIÓN 14. PANEL PARA AÑADIR UN PERIÓDICO | 37 |
| ILUSTRACIÓN 15. PANEL PARA MODIFICAR UN PERIÓDICO..... | 38 |
| ILUSTRACIÓN 16. PANEL PARA BORRAR UN PERIÓDICO..... | 38 |
| ILUSTRACIÓN 17. PANEL PARA AÑADIR NOTICIAS | 39 |
| ILUSTRACIÓN 18. PANEL PARA VISUALIZAR DATOS Y RESULTADOS | 39 |
| ILUSTRACIÓN 19. PANEL PARA AÑADIR NOTICIAS MEDIANTE BÚSQUEDA..... | 45 |
| ILUSTRACIÓN 20. DATOS DE NOTICIA DEL PERIÓDICO ONLINE 20 MINUTOS..... | 46 |
| ILUSTRACIÓN 21. DATOS DE COMENTARIO DEL PERIÓDICO 20 MINUTOS | 47 |
| ILUSTRACIÓN 22. DATOS DE NOTICIA DEL PERIÓDICO ONLINE EL MUNDO | 47 |
| ILUSTRACIÓN 23. DATOS DE COMENTARIO DEL PERIÓDICO EL MUNDO..... | 48 |
| ILUSTRACIÓN 24. AÑADIR UN PERIÓDICO | 62 |
| ILUSTRACIÓN 25. EJEMPLO DE URL BASE PARA EL PERIÓDICO 20 MINUTOS..... | 62 |
| ILUSTRACIÓN 26. URL DE BÚSQUEDA | 63 |
| ILUSTRACIÓN 27. MODIFICAR UN PERIÓDICO | 64 |
| ILUSTRACIÓN 28. BORRAR UN PERIÓDICO | 65 |
| ILUSTRACIÓN 29. AÑADIR NOTICIAS | 66 |
| ILUSTRACIÓN 30. PANEL DE VISUALIZACIÓN | 67 |
| ILUSTRACIÓN 31. GUARDADO DE IMÁGENES DE RESULTADOS | 68 |

Glosario

- HTML (HyperText Markup Language): Lenguaje para la creación de páginas web.
- XML (Extensible Markup Language): Lenguaje para almacenar información de forma legible basado en etiquetas.
- URL (Uniform Resource Locator): Identificador de recursos uniforme, o comúnmente llamadas enlaces de páginas web.
- RSS: (Really Simple Syndication): Sindicación realmente simple. Se trata de un canal web para difundir información.
- NetBeans: Entorno de desarrollo de software.
- Jsoup: Librería para Java para analizar, descargar y obtener el contenido de los elementos del código HTML.
- PhantomJS: Aplicación de tipo navegador sin interfaz, o headless browser.
- CasperJS: Aplicación que utiliza PhantomJS para ser capaz de navegar por páginas web mediante comandos y obtener su contenido.
- DOM (Document Object Model): Modelo de Objetos del Documento. Modelo del documento disponible en navegadores para tener acceso a los elementos del código HTML.
- XStream: Librería de Java que sirve para serializar y recuperar objetos en el formato XML.
- JFreeChart: Librería para Java que sirve para realizar diagramas.
- BBDD: Base de datos
- Script: Fichero de código, normalmente de pequeño tamaño, utilizado para realizar una tarea específica.
- PNG: Portable
- SVG, (Scalable Vector Graphics): Fichero de gráficos vectoriales y escalables.
- PDF, (Portable Document File), formato para guardar documentos.
- JSON: (JavaScript Object Notation), Notación de objetos de JavaScript. Sirve para almacenar y e intercambiar datos.

1 Introducción

1.1 Motivación

A diario se difunden grandes cantidades de información por diferentes medios online. Un mismo hecho puede ser narrado de diferentes formas: de manera lo más objetiva posible, de manera parcial intencionadamente, o incluso de manera sesgada sin tener la intención de hacerlo así. La manera en que se escribe la información puede influir en la percepción que los lectores tienen de la misma, incluso independientemente del hecho narrado.

Por ello resultaría muy interesante y útil poder analizar una misma información publicada en distintos medios online, así como la reacción y las emociones que suscita a los lectores de cada uno de ellos.

Para ello sería necesario extraer y almacenar esta información, realizar un análisis de las emociones que transmiten y poder visualizar tanto la información como el resultado del análisis de emociones de manera gráfica y accesible, con el fin de facilitar su posterior estudio.

1.2 Objetivos

Este trabajo tiene como objetivo principal dar soporte a la extracción de noticias de periódicos online de forma lo más general posible y facilitar la representación y visualización tanto de la información obtenida como del resultado del análisis de emociones.

De este objetivo se derivan los siguientes objetivos específicos para la obtención, análisis de emociones y representación de la información:

- Facilitar la incorporación, modificación y eliminación de métodos de extracción de noticias.
- Obtener la información de las noticias de diferentes periódicos online
- Obtener la información de los comentarios de las noticias.
- Analizar emocional y sentimentalmente los textos de las noticias y comentarios.
- Incorporar una interfaz para visualizar los datos y los resultados del análisis de emociones de las noticias y comentarios almacenados.
- Facilitar la comparación de la información de los textos con el análisis de emociones de los mismos.
- Almacenar la información de las noticias y comentarios, así como los resultados del análisis emocional.
- Recuperar y volver a visualizar información extraída y analizada previamente.

1.3 Estructura del documento

En este apartado se detallan las partes en las que está estructurado este documento.

Sección 1: Introducción. En esta sección se describe la motivación del proyecto y sus objetivos. También se describe la estructura del documento en sí.

Sección 2: Estado del arte. En este apartado se encuentran investigaciones y trabajos relacionados con el proyecto realizado.

Sección 3: Análisis. Descripción del problema para conocer sus características y enumeración de requisitos del sistema.

Sección 4: Diseño. Descripción de la arquitectura del sistema y cómo se ha estructurado, así como el detalle de la extracción dinámica de datos y del análisis de emociones.

Sección 5: Desarrollo. Descripción de cómo se han implementado los distintos módulos del diseño.

Sección 6: Ejemplos. Ejemplos de uso de la herramienta y comparación de noticias extraídas y resultados.

Sección 7: Pruebas y resultados. Enumeración de test realizados a cada módulo y resultados.

Sección 8: Manual de usuario: Manual práctico para utilizar cada una de las funciones disponibles del sistema.

Sección 9: Conclusiones. Presentación de conclusiones, ventajas, limitaciones de la herramienta. Sugerencias de aprovechamiento de uso futuro y consideraciones finales.

Sección 10: Referencias. Enumeración de referencias utilizadas para la realización del trabajo.

2 Estado del arte

En este apartado se describen algunos trabajos relacionados con el proyecto:

En el contexto del análisis de las noticias, en algunos trabajos como [1] se estudia la relación de la confianza hacia los políticos y el uso de medios online mediante un análisis de sentimiento entre sitios web de información y redes sociales, concluyendo que los usuarios de los sitios web están vinculados a una mayor confianza en los políticos que los usuarios de redes sociales. En otro artículo [2] se presenta un sistema que, mediante técnicas de análisis sentimental, asigna a cada palabra un valor positivo o negativo. Para incrementar la fiabilidad en el resultado final se agrega un sinónimo y un antónimo, realizando esta acción durante varias iteraciones, asignándoles también un valor positivo o negativo, siendo los antónimos el valor contrario. En aquellos caminos que se desvíe menos la polaridad en las palabras se tiene más confianza del valor de la polaridad. Posteriormente utilizan este sistema con el cuerpo de los textos en noticias y blogs que hablan sobre personas famosas, con el fin de conocer la puntuación total con el número de palabras positivas o negativas.

Continuando con los trabajos relacionados con el análisis sentimental, en [3] se realiza un sistema que analiza comentarios de reseñas de hoteles y ofrece una visión de conjunto de las reseñas, resumiendo así las opiniones para un determinado hotel. En [4] se plantea un enfoque para el análisis de sentimiento en textos con oraciones compuestas a nivel de frases, para determinar si una expresión es neutral o tiene alguna polaridad, para después obtener el valor de la polaridad, negativa o positiva.

En el ámbito del análisis de sentimientos o emociones en los medios online y las redes sociales, en [5] se analizan los mensajes o textos que los usuarios escriben en la red social Facebook para determinar la polaridad de estos: positiva, neutra o negativa, permitiendo la comparación entre el estado emocional “habitual” de cada usuario y posibles estados emocionales puntuales, que podrían ser indicadores de algún cambio a ser considerado.

En el contexto del análisis de textos en el artículo [6] se investiga la detección automática de la ironía y el humor en redes sociales enfocándolo como un problema de clasificación.

Para finalizar, en el trabajo de fin de grado [7] se creó una herramienta que sirvió de primera aproximación hacia la extracción y análisis de emociones de los textos de noticias y comentarios de un periódico online concreto, basado en la comparación de las palabras de los textos con las recogidas en los diccionarios de emociones utilizados. En este trabajo se trata de dar un soporte más amplio a la extracción de información de distintas fuentes y al análisis de la misma, creando una plataforma de carácter general que permite integrar noticias y comentarios de distintos periódicos, así como realizar un análisis de las emociones que transmiten, y de ofrecer también soporte a la integración, visualización y comparación de los distintos elementos y resultados del análisis llevado a cabo.

3 Análisis

Como se ha mencionado anteriormente, en este trabajo no solo se dará soporte a la extracción de noticias y comentarios de lectores de periódicos online concretos, y al análisis de las emociones que transmiten. También se desarrollará una herramienta que permita obtener e integrar la información de las noticias de distintos periódicos online, y que permita visualizar los datos extraídos y el resultado del análisis de las emociones que transmiten. Además, es necesario poder añadir nuevos periódicos de manera dinámica, sin necesidad de modificar el código de la propia herramienta. Se precisa, por lo tanto, dotar a esta nueva herramienta de una interfaz gráfica que permita la realización de las tareas anteriores de manera sencilla y eficaz.

3.1 Descarga de información de los periódicos online

Se comenzó usando la herramienta anterior para intentar aclarar cómo incorporar las nuevas funcionalidades. Para ello se intentó descargar las noticias del periódico 20 Minutos que estaban programadas para descargar cuando se ejecutara, siendo estas noticias antiguas, así como utilizando la descarga de las noticias en los RSS, siendo estas otras actuales. La prueba resultó fallida, puesto que no se fue capaz de obtener ninguna noticia.

Tras analizar el comportamiento de la aplicación se vio que en el código HTML resultante de la descarga de las noticias había una zona donde deberían aparecer los comentarios. Sin Se utilizó como punto de partida el software desarrollado durante el trabajo fin de grado [6] y se comprobó que ya no funcionaba ni para la descarga de las noticias concretas del periódico 20 Minutos ni para la descarga de las noticias actuales de los RSS. Por otra parte, el sistema de comentarios anterior había sido reemplazado por uno nuevo. El software de descarga desarrollado previamente sólo era capaz de obtener el código HTML estático de la página, siendo imposible obtener ni las noticias ni los comentarios en la actualidad.

También se analizó la forma de descargarse noticias de otros periódicos online, como El País, El Mundo o La Vanguardia. En cada uno de ellos se utiliza código dinámico en JavaScript para generar el código y mostrar el contenido de la noticia posteriormente, después de la carga inicial de la página. Es más, en algunos periódicos como El Mundo, los comentarios no se cargan automáticamente. Aparece un botón flotante a la derecha de la noticia, siguiendo los movimientos con el texto «Ver comentarios». Debajo de la noticia, si la noticia dispone de ellos, se muestra un comentario únicamente y debajo otro botón para mostrar los comentarios. Al pulsar cualquiera de los dos botones aparece un panel desde el lado derecho del navegador con la sección de comentarios. Para periódicos como este, además de poder ejecutar JavaScript durante la obtención de noticias, sería necesario tener la capacidad de pulsar botones dentro de la página.

Tras este análisis inicial, se concluyó que el método de descarga de noticias debería tener soporte para JavaScript y para la realización de ciertas acciones antes de descargar el código HTML resultante, como por ejemplo pulsar botones de la página o mover el desplazamiento de la página al final para cargar todo el contenido de esta.

3.2 Extracción de información de manera genérica

Para obtención la información concreta de las noticias de los periódicos online es necesario analizar la estructura del código HTML que genera cada periódico. Como las noticias de cada uno tienen su propia estructura (distinta según el periódico), es necesario especializar la extracción de información para cada uno de ellos.

Tras investigar las diferentes maneras de obtener la información, y habiendo visto que incluso dentro del mismo periódico online el formato de las noticias puede cambiar en cualquier momento, es claro que es necesario especializar la extracción de información para cada caso.

El hecho de que tanto la forma de descargar noticias como la estructura de las mismas y de los comentarios asociados sean diferentes para cada caso, confirmó la necesidad de crear herramientas genéricas que permitan dar soporte a la extracción y análisis de noticias, de forma que cuando se desee cambiar de periódico o se produzcan actualizaciones en alguno de ellos solo haya que cambiar un script situado en una herramienta adicional externa, sin que haya que modificar la propia herramienta de análisis de noticias en sí. Para que todo funcione correctamente, la herramienta externa de descarga y extracción de noticias y comentarios deberá ser llamada desde la herramienta principal. Será la información de salida de la herramienta externa la que deberá mantener un formato estandarizado, de modo que la herramienta principal pueda reconocer la incorporar la información a la plataforma. La Ilustración 1 muestra un esquema general de esta propuesta. Más adelante, en la Ilustración 4, se mostrará cómo se integra esta parte de extracción de información en la plataforma completa para la visualización, comparación y análisis de noticias, comentarios y emociones que transmiten.

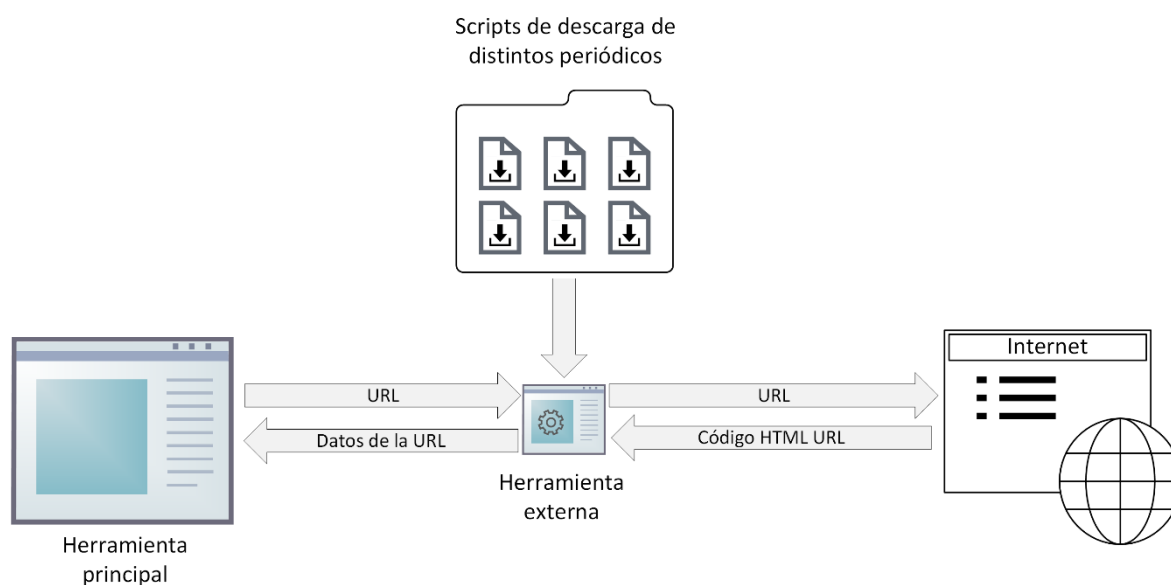


Ilustración 1. Esquema de extracción de información de manera genérica

3.3 Métodos de obtención de noticias

Con el fin de maximizar las posibilidades de obtención de noticias de cada periódico, durante el análisis se han barajado varias opciones o métodos de extracción de información.

El primero es el de obtener las noticias mediante su URL. Es el método más básico, ya implementado para un periódico concreto anteriormente. En este caso, al trabajar con distintos periódicos, el método deberá ser algo más sofisticado. Será necesario almacenar, para cada periódico, un campo de datos con información sobre la URL base de sus noticias, es decir, la parte común con la que empiezan todas las URL de las noticias de ese periódico. Comparando la URL de la noticia a descargar con todas las URL base de los periódicos se podrá saber a qué periódico pertenece la noticia y, de esta manera, realizar la descarga del contenido de la noticia de ese periódico concreto.

El segundo método, también contemplado previamente de forma específica para un periódico en concreto, consiste en descargar las noticias de los RSS del periódico. También en este caso la implementación deberá ser más sofisticada, al disponer de varios periódicos desde donde descargar el contenido, cada uno con su RSS. Se deberá poder dar a elegir al usuario de qué periódico descargar las últimas noticias que aparecen en los RSS. También deberá poderse dar la opción de descargar todas las noticias de los RSS de todos los periódicos a la vez.

El último método de descarga se basará en una búsqueda de noticias relacionadas con las palabras clave que se indiquen. Este nuevo método permitirá obtener varias noticias que tengan relación con las palabras clave introducidas. También permitirá elegir si la búsqueda será realizada solo en un periódico seleccionado o en todos a la vez. Esto es muy útil para facilitar que, a partir de una única búsqueda, se puedan comparar noticias sobre el mismo tema dentro del mismo periódico o en todos los periódicos contemplados en la plataforma, sin que sea necesario buscar cada una de las noticias de esa temática en cada uno de los periódicos en internet, y obtener sus URL e introducirlas a mano en la herramienta.

Para que esta funcionalidad se pueda incorporar a la herramienta se deben dar dos condiciones en el periódico online:

- El periódico online debe tener la funcionalidad de realizar búsqueda de noticias.
- Las búsquedas en el periódico online deben realizarse utilizando el método GET de HTML.

La Ilustración 2 muestra la URL resultante tras realizar una búsqueda en el periódico online 20 Minutos. El método GET se caracteriza por utilizar los parámetros de la llamada en la URL. Este periódico cumple ambas condiciones para utilizar el método de obtención de noticias mediante búsqueda.

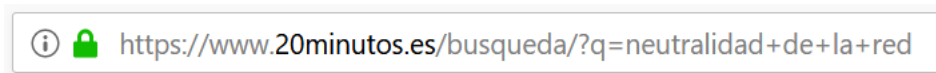


Ilustración 2. URL resultante tras realizar una búsqueda en el periódico online 20 Minutos

Para dar soporte a este método de extracción de noticias, para cada periódico se deberá almacenar un campo adicional de datos que contenga la URL base para realizar búsquedas, a la que se le añadirán las palabras clave que introduzca el usuario para obtener las noticias relacionadas con esas palabras clave.

3.4 Funcionalidad y visualización de información

Para facilitar la visualización y comparación de las noticias de distintos periódicos, los comentarios asociados y las emociones que se detectan en cada caso, es necesario incorporar a la herramienta una interfaz de usuario. Esta interfaz debe dar soporte a las siguientes funcionalidades clave:

- Visualización de información de periódicos, noticias y comentarios extraídos y las emociones que transmiten.
- Visualización de la representación gráfica de los resultados del análisis de las emociones detectadas en los diferentes periódicos, noticias y comentarios.
- Gestión de periódicos en la plataforma, incluyendo adición, modificación y eliminación.
- Obtención de noticias desde URL, desde RSS y desde búsqueda.
- Almacenamiento en la base de datos de información sobre los periódicos, las noticias, los comentarios y el análisis de emociones realizado a sus textos.

Para ofrecer una correcta visualización de la información se necesita un mecanismo que permita navegar entre toda la información disponible. La información estará distribuida en torno a un conjunto de periódicos, cada periódico tendrá un conjunto de noticias y cada noticia tendrá un conjunto de comentarios. El usuario podrá seleccionar un periódico entre todos los disponibles. Al seleccionar uno, la herramienta mostrará la lista completa de noticias de este periódico, permitiendo seleccionar cualquiera de ellas. Por último, se mostrarán los comentarios asociados a esa noticia, que se podrán seleccionar de forma individual. También sería conveniente dar la opción de seleccionar todos los elementos de cada lista. La herramienta representará y permitirá la visualización de:

- Información de todos los periódicos: Cuando se ha seleccionado la opción de todos los periódicos en la lista de selección de periódicos.
- Información de un periódico y todas sus noticias: Cuando se ha seleccionado un periódico de la lista, pero aún no se ha seleccionado ninguna noticia suya o se ha seleccionado la opción de todas las noticias.
- Información de un periódico, una de sus noticias y todos sus comentarios: Cuando se haya seleccionado un periódico y una noticia, sin haber seleccionado ningún comentario aún o habiendo seleccionado la opción de ver todos los comentarios.
- Información de un periódico, una de sus noticias y uno de sus comentarios: Cuando se haya seleccionado un periódico, una noticia y un comentario.

Con ello se puede afinar la búsqueda de información, pudiendo visualizar toda la información de la base de datos al completo, incluyendo los resultados del análisis de emociones y la representación gráfica de cada uno de ellos, pudiendo comparar los resultados obtenidos para las distintas noticias y periódicos. La interfaz permitirá realizar la extracción y consulta de información de manera sencilla, rápida e intuitiva, sin necesidad de que la herramienta sea modificada al incorporar nuevos periódicos o formas de extraer información.

3.5 Análisis de emociones

Para realizar el análisis de emociones se utilizarán cuatro diccionarios de palabras que representan las emociones Joy (Alegría), Anger (Ira), Fear (Miedo) y Sadness (Tristeza).

Cada diccionario de emoción consta de tres tipos de palabras: palabras “tal cual”, que serán comparadas tal y como están escritas, incluyendo su plural; palabras “sufijos”, que serán comparadas incluyendo todos los posibles sufijos que puedan tener; y palabras “verbos”, que serán comparadas con todas las formas temporales, de género y número de la propia conjugación que tenga ese verbo. Si una palabra del texto a analizar coincide pasará a formar parte de una lista. El número de palabras encontradas que pertenecen a una lista de emoción se trata del número de esas emociones encontradas en el texto.

Este método fue explorado durante la realización del trabajo fin de grado. Sin embargo, en ocasiones se vio que los textos analizados no detectaban correctamente las emociones. Esto es debido a que en una frase se le puede dar un sentido a una expresión que induzca a que tenga el sentido contrario, por ejemplo, la negación, o algo mucho más complejo de detectar como la figura retórica de la ironía.

Para realizar un análisis más completo de los textos extraídos se ha incluido un indicador adicional, pudiendo ser de ayuda para determinar las emociones reales subyacentes del texto. Este analizador adicional permite identificar la polaridad o sentido de un texto en términos de si tiende a transmitir emociones positivas, neutras o negativas [8].

El hecho de que la herramienta permita analizar las emociones transmitidas junto con la polaridad de los mensajes, y que además permita visualizar los textos y el resultado de su análisis, facilita la detección de falsos positivos durante el uso habitual de la herramienta, y no sólo durante su desarrollo y la realización de las pruebas. Esto permite retroalimentar el sistema, que podrá beneficiarse de las observaciones realizadas durante la interacción con la herramienta.

3.6 Representación de emociones

Para la representación del análisis de emociones se necesita una herramienta que sea capaz de generar gráficos para representar las emociones de diferentes maneras y guardarlos en ficheros de imagen.

Por otro lado, se necesitan mostrar las imágenes de la representación de emociones en una interfaz gráfica, por lo que se necesita una herramienta que además cumpla con esta característica.

3.7 Almacenamiento de datos

Para almacenar los datos de la aplicación en tiempo de ejecución se utilizarán los objetos de las clases destinadas a mantener los datos de la herramienta.

Se necesita una herramienta con permita guardar los datos de los periódicos, las noticias y los comentarios, así como el análisis de emociones de los textos en un fichero. Además, esta herramienta ha de ser capaz de leer este fichero de nuevo y de manera inversa restaurar la información de los objetos de datos de la herramienta.

3.8 Requisitos

A continuación, se muestran los requisitos educidos como resultado de la unión y clasificación de todas las necesidades extraídas durante el análisis realizado y descrito previamente.

3.8.1 Requisitos funcionales

3.8.1.1 Requisitos de extracción de información de noticias

La herramienta permitirá:

EXT-01: Obtener la información de noticias de un periódico online.

EXT-02: Obtener la información de noticias desde distintos periódicos online de manera simultánea.

EXT-03: Obtener la información de una noticia a partir de la URL de la noticia.

EXT-04: Detectar a qué periódico pertenece cada URL introducida para proceder a descargar su información correctamente.

EXT-05: Obtener la información de las noticias de un periódico a partir de la URL de los RSS del periódico si está disponible en el periódico.

EXT-06: Elegir al usuario desde qué periódico se obtendrán las noticias desde los RSS.

EXT-07: Obtener al usuario las noticias desde los RSS de todos los periódicos a la vez.

EXT-08: Obtener información de noticias de un periódico a partir de una búsqueda con palabras clave si está disponible en el periódico.

EXT-09: Elegir al usuario en qué periódico realizar la búsqueda de noticias con palabras clave.

EXT-10: Obtener al usuario las noticias usando palabras clave de todos los periódicos a la vez.

EXT-11: Obtener la siguiente información de una noticia: URL, titular, texto, autor, fecha, votos positivos y negativos, y comentarios de la noticia, siempre y cuando estos datos estén disponibles.

EXT-12: obtener la siguiente información de un comentario: autor, fecha, votos positivos y negativos y texto del comentario, siempre y cuando estos datos estén disponibles.

3.8.1.2 Requisitos de gestión de periódicos genéricos

GEP-01: La información necesaria mínima al añadir un periódico será: nombre URL base y script de descarga.

GEP-02: La herramienta deberá permitir al añadir un periódico incorporar la información necesaria para obtener noticias a través de los RSS.

GEP-03: La información de los RSS al añadir un periódico es: URL de los RSS y script de descarga de RSS.

GEP-04: La información de los RSS al añadir un periódico será opcional.

GEP-05: La herramienta deberá permitir al añadir un periódico incorporar la información necesaria para obtener noticias a través de la búsqueda.

GEP-06: La información de la búsqueda al añadir un periódico es: URL de búsqueda y script de descarga de búsqueda.

GEP-07: La información de búsqueda al añadir un periódico será opcional.

GEP-08: La herramienta no debe permitir introducir información de RSS incompleta al añadir un periódico.

GEP-09: La herramienta no debe permitir introducir información de búsqueda incompleta al añadir un periódico.

GEP-10: La información necesaria mínima al modificar un periódico será: nombre URL base y script de descarga.

GEP-11: La herramienta deberá permitir al modificar un periódico incorporar la información necesaria para obtener noticias a través de los RSS.

GEP-12: La información de los RSS al modificar un periódico es: URL de los RSS y script de descarga de RSS.

GEP-13: La información de los RSS al modificar un periódico será opcional.

GEP-14: La herramienta deberá permitir al modificar un periódico incorporar la información necesaria para obtener noticias a través de la búsqueda.

GEP-15: La información de la búsqueda al modificar un periódico es: URL de búsqueda y script de descarga de búsqueda.

GEP-16: La información de búsqueda al añadir un periódico será opcional.

GEP-17: La herramienta no debe permitir introducir información de RSS incompleta al modificar un periódico.

GEP-18: La herramienta no debe permitir introducir información de búsqueda incompleta al modificar un periódico.

La herramienta debe permitir borrar periódicos introducidos por el usuario.

La herramienta permitirá al usuario elegir qué periódico eliminar.

GEP-19: El sistema no podrá tener periódicos con el mismo nombre.

3.8.1.3 Requisitos de almacenamiento y datos:

DAT-01: Los datos que compondrán un periódico serán: nombre, URL base, script de descarga de noticia, URL de RSS, script de descarga de RSS, URL de búsqueda, script de descarga de búsqueda y una lista de noticias.

DAT-02: Los datos que compondrán una noticia serán: titular de la noticia, URL de la noticia, texto de la noticia, autor de la noticia, votos positivos, votos negativos, fecha, emoción de la noticia, periódico al que pertenece la noticia y una lista de comentarios.

DAT-03: Los datos que compondrán un comentario serán: texto del comentario, autor del comentario, fecha, votos positivos, votos negativos,

DAT-04: Los datos que compondrán una emoción serán: texto, lista de palabras separadas, lista de palabras de cada emoción y el sentimiento.

DAT-05: Los scripts de descarga de los distintos tipos de extracción deben estar escritos para ejecutarse con CasperJS.

DAT-06: Los scripts de descarga deben poder soportar los siguientes parámetros de entrada: URL de descarga, nombre de fichero de salida.

DAT-07: El script de descarga por URL debe guardar un fichero con toda la información extraída de la noticia (DAT-02).

DAT-08: El script de descarga por RSS y por búsqueda debe guardar un fichero con una lista de URL.

DAT-09: El sistema debe poder leer de un fichero los datos de los periódicos y noticias descargadas al iniciar la aplicación.

DAT-10: La herramienta permitirá guardar los datos de los periódicos y las noticias descargadas en un fichero.

3.8.1.4 Requisitos de Interfaz

GUI-01: El sistema dispondrá de una ventana gráfica desde la cual el usuario pueda realizar las distintas acciones.

GUI-02: La herramienta dispondrá de un panel para cada una de las siguientes funcionalidades: añadir un periódico, modificar un periódico, borrar un periódico, añadir noticias, visualizar datos.

GUI-03: La herramienta tendrá un menú desde el que se podrán acceder a los paneles y acciones disponibles.

GUI-04: La interfaz deberá permitir acceder a cualquier panel en cualquier momento.

GUI-05: La interfaz debe permitir realizar una copia de la información de los periódicos y las noticias descargadas a un fichero.

GUI-06: La interfaz debe permitir al usuario seleccionar los datos a ser mostrados; selección de periódicos, selección de noticias y selección de comentarios.

GUI-07: La interfaz deberá mostrar la información de la selección de los periódicos, noticias y comentarios seleccionados.

GUI-08: La interfaz deberá mostrar una representación de emociones encontradas en los textos en función de los periódicos, noticias y comentarios seleccionados.

GUI-09: La interfaz deberá mostrar al usuario un mensaje de la correcta o errónea ejecución de las acciones emprendidas, entre las cuales serán: añadir un periódico, modificar un periódico, eliminar un periódico, número de noticias obtenidas.

3.8.1.5 Requisitos de análisis de emociones

ANE-01: La herramienta realizará un análisis de emociones de los textos de las noticias y comentarios descargados.

ANE-02: El análisis de emociones encontrará las palabras del texto que pertenezcan a una determinada emoción y se almacenarán, para cada emoción, en una lista.

ANE-03: El análisis del sentimiento dará una puntuación al texto de entre 0 y 1 en función si el texto tiene connotaciones negativas o positivas respectivamente.

3.8.2 Requisitos no funcionales

3.8.2.1 Requisitos de rendimiento y estabilidad

El sistema ha de ejecutarse de manera eficiente y evitando quedarse a la espera en operaciones cuando el resultado de la ejecución no es el esperado. En caso de errores o fallo el sistema debe ser capaz de recuperarse de manera controlada y seguir adelante, informando al usuario pertinentemente.

3.8.2.2 Requisitos de usabilidad

El sistema debe mostrar la información al usuario de manera clara, utilizando las etiquetas identificativas necesarias y permitiendo el uso intuitivo de cada función. También deben ocultarse funciones que no estén permitidas pertinentemente y evitando que el usuario introduzca datos erróneos.

3.8.2.3 Requisitos de mantenibilidad

El código de la herramienta ha de ser claro, modular y escalable, para permitir la ampliación, modificación o eliminación de funcionalidad de manera sencilla y sin que afecte al resto del sistema.

3.9 Tecnologías a utilizar

3.9.1 Lenguaje y entorno de programación

Para el desarrollo de este trabajo se ha planteado utilizar uno de los que probablemente sean los dos lenguajes de programación más extendidos, los lenguajes C y Java.

El lenguaje C es estructurado y su uso está más orientado hacia la programación de bajo nivel. Permite libertad a la hora de establecer tipos de datos, pero también es más costoso.

Java es un lenguaje orientado a objetos. Cuenta con la ventaja de una mayor facilidad para definir y manejar estructuras de datos utilizando las clases.

Teniendo en cuenta las estructuras de datos analizadas en los capítulos anteriores se ha decidido realizar el proyecto utilizando Java. También se ha tenido en cuenta la experiencia contraída con el uso de Java pudiendo sacar más provecho de este lenguaje. Además, una de las utilidades del lenguaje Java que se pueden aplicar en esta herramienta será la herencia de clases. Esto permitirá al definir y utilizar nuevos tipos de periódicos reutilizar código y tener una estructura más coherente y realista.

3.9.2 Descarga de información de páginas web

Como se ha visto en la sección 3.1 Descarga de información de los periódicos online, se necesita un programa que sea capaz de descargar el código HTML de las páginas web de las noticias.

Las páginas web de los periódicos online suelen utilizar pequeños fragmentos de código JavaScript para cargar el contenido de las páginas de manera dinámica. Utilizando Jsoup sólo se consigue descargar el contenido estático. Otras páginas requieren de acciones, como pulsar botones, para que este contenido dinámico sea activado. Por ello se requiere una utilidad que sea capaz de trabajar con JavaScript y permita navegar por las páginas web antes de descargar el código HTML final con todo el contenido.

Existen actualmente unas aplicaciones llamadas *headless browser* o navegadores sin interfaz gráfica. Estas aplicaciones interpretan la página web de manera similar a lo que lo haría un navegador web convencional, sin utilizar una interfaz gráfica. El propósito de la mayoría de estas herramientas es realizar tests automatizados sobre las páginas web. Eso significa que se intenta reproducir un escenario lo más realista posible. Es decir, están preparados para ejecutar el código JavaScript para poder cargar todo el contenido dinámico y también realizar las operaciones necesarias para mostrar otras partes de las páginas.

Entre todas las herramientas de navegadores sin cabeceras disponibles se han preseleccionado las siguientes candidatas: Htmlunit y CasperJS.

Htmlunit es una librería para Java que simula un navegador web, pensado para realizar tests unitarios. Es capaz de cargar páginas web como un navegador corriente lo haría actualmente y permite navegar por la página realizando diferentes acciones como pulsar botones o rellenar y enviar formularios. También permite, utilizando unas funciones específicas, acceder a cualquier elemento del código directamente, por ejemplo, el título de la noticia.

CasperJS es una aplicación en forma de capa de software que utiliza el navegador sin interfaz PhantomJS. Permite cargar páginas web completas, como un navegador convencional, ya que soporta JavaScript. Al estar este software orientado a realizar test sobre páginas web permite realizar acciones sobre la web, pudiendo obtener el código de la página web tras pulsar y ejecutar todas las partes de código dinámicas.

Este software se ejecuta como una aplicación independiente, no formando parte de una librería Java. Esto añade la funcionalidad de utilizar ficheros de scripts externos a la herramienta principal en los cuales se pueden crear escenarios de navegación específicos. El lenguaje utilizado en los scripts es JavaScript junto a funciones específicas para el manejo de CasperJS. Al utilizar JavaScript, se tienen disponibles los métodos HTML DOM, que sirven para acceder a los elementos del código HTML una vez descargado.

De las dos candidatas, la herramienta que mejor se ajusta a las necesidades es CasperJS. Aunque las dos cubren prácticamente la totalidad de la funcionalidad requerida, es CasperJS la elegida debido a que permite implementar múltiples ficheros de script externos que realicen las tareas de descarga de distinta manera, permitiendo que se ajuste cada uno a un periódico distinto. De esta manera la herramienta principal no se ve afectada por los posibles cambios en la estructura del HTML de las noticias, pudiendo modificar el script externo que falle o eliminarlo, manteniendo la integridad de la aplicación.

3.9.3 Analizador de código HTML

Como tarea posterior a la descarga del código HTML se necesita un analizador de código capaz de extraer los elementos necesarios de él. Como se ha visto en los apartados anteriores, Jsoup ya no sirve como herramienta de descarga de noticias, al no ser estas páginas HTML estáticas. Sin embargo, sigue sirviendo para analizar y extraer la información del código HTML una vez descargado. Al ser una librería de java el código que se utilice para extraer la información formará a ser parte de la herramienta principal y, si alguna vez cambia la estructura del periódico habrá que modificarlo.

Por otro lado, la nueva herramienta para descargar código HTML, CasperJS, está basada en JavaScript y también sirve para analizarlo y extraer elementos específicos de él. El hecho de que se trate de una herramienta externa y que se puedan utilizar diferentes scripts para descargar hace que sea el más idóneo para obtener los periódicos genéricos. Sin embargo, el uso de CasperJS para extraer la información es algo más complejo, al no estar integrado en un entorno en el que se pueda depurar.

Para no cerrar ninguna puerta a futuras ampliaciones se ha decidido, por un lado, mantener la librería Jsoup para realizar el análisis del código HTML para los periódicos embebidos en la herramienta e incorporar CasperJS para realizar el análisis del código HTML y la extracción de datos para los periódicos genéricos de manera externa.

3.9.4 Interfaz gráfica

Las interfaces gráficas de usuario que se han analizado han sido JavaFX y Swing.

Swing es una librería de interfaz gráfica para Java. Está ideada para ser, al igual que Java, compatible sin importar mucho en donde se ejecute. Dispone de muchos elementos para ser utilizados en interfaces como paneles, botones, cajas de texto, etc.

JavaFX es una plataforma de software pensada inicialmente para aplicaciones web. Las aplicaciones JavaFX son compiladas a Java, por lo que se pueden ejecutar en la misma máquina virtual. También se puede utilizar código Java en aplicaciones de JavaFX. Esta plataforma también tiene componentes gráficos muy potentes y avanzados como formas animadas y efectos visuales.

A pesar de ser JavaFX potencialmente más avanzada, no se necesitan componentes tan avanzados para la realización de este trabajo. Se hicieron pruebas de uso y finalmente se decidió por realizar la interfaz gráfica utilizando Swing al primar la sencillez en el diseño y funcionalidad.

3.9.5 Almacenamiento de datos

XStream es una librería para java que se utiliza para serializar objetos en formato XML. Una vez serializados, estos datos pueden ser fácilmente guardados en un fichero, siendo posible realizar el sentido inverso para recuperar datos desde un fichero hacia la información de la aplicación.

Por su sencilla forma de uso se utilizará esta utilidad, XStream, para almacenar la información de la aplicación. Se guardarán también los nuevos objetos que resultarán al incorporar los periódicos genéricos.

3.9.6 Análisis de emociones

Existe un servicio en la nube llamado Cognitive Services de la empresa Microsoft que brindan la posibilidad de incorporar a aplicaciones o sitios webs diferentes algoritmos. En este servicio hay una sección dedicada al procesamiento del lenguaje natural, evaluación de sentimientos y reconocimiento del lenguaje.

Uno de los servicios ofrecidos es el análisis del sentimiento de textos, el cual tras analizarlos devuelve un valor entre 1 y 0, que indica si el texto tiene una impresión positiva o negativa respectivamente. Valores cercanos a 0.5 indican neutralidad. El texto a analizar se puede enviar a los servidores, los cuales utilizan técnicas de aprendizaje automático, con modelos de datos entrenados con este fin.

A pesar de ser un servicio de pago en su uso a gran escala, existe un nivel gratuito con limitaciones en el que puede utilizarse hasta 5.000 llamadas por mes, por lo que aun así se podrían analizar más de 100 noticias por día.

Asimismo, utilizar este indicador en los textos puede ser muy útil a la hora de verificar los resultados de otros análisis de emociones realizados a los textos. Por estos motivos se ha decidido incorporar para el análisis de emociones en la herramienta.

3.9.7 Representación de emociones

Para representar las emociones se necesita una herramienta que sea capaz de mostrar en los paneles de la interfaz gráfica los diagramas de representación del resultado del análisis de las emociones correspondientes.

Tras realizar algunas pruebas, se ha decidido utilizar la herramienta JFreeChart, debido a que permite realizar diagramas de tipo tarta, entre otros, y es compatible con los paneles de la interfaz gráfica a utilizar.

4 Diseño

4.1 Arquitectura general del sistema

El sistema se ha diseñado siguiendo el patrón «Modelo Vista Controlador», es decir, separando los datos de la herramienta (Modelo), de la interfaz gráfica (Vista) y de la lógica de la aplicación (Controlador).

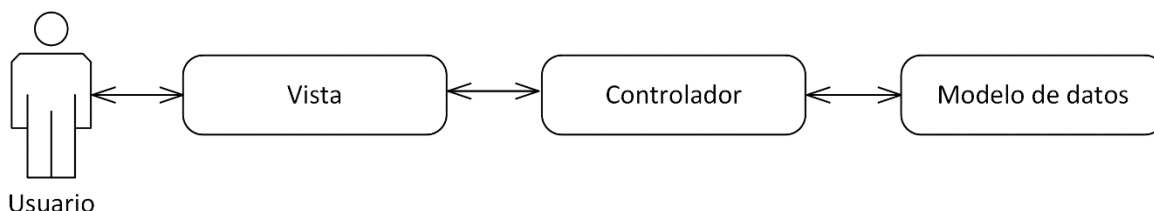


Ilustración 3. Esquema de Modelo-Vista-Controlador

Como se puede observar en la Ilustración 4, el usuario tiene disponible una serie de funcionalidades que puede ejecutar a través de la interfaz gráfica: obtener noticias con una URL, obtener las noticias de los RSS, obtener noticias por medio de búsqueda, crear, modificar y eliminar periódicos, guardar la base de datos del sistema y visualizar la información de los datos descargados y las emociones que transmiten según el análisis de emociones realizado.

El usuario también puede crear o modificar los scripts correspondientes a los periódicos para obtener los datos de noticias de la fuente de información que el usuario desee. Estos scripts se mantendrán en el directorio raíz de la aplicación y se podrán utilizar en la misma utilizando las funcionalidades de *añadir periódico* o *modificar periódico*.

Como consecuencia de la interacción del usuario con la interfaz se ejecutarán las acciones asociadas a cada funcionalidad a la que se da soporte en la parte del controlador. El controlador se encarga de comunicarse con la herramienta CasperJS para descargar las noticias, así como también de procesar esos datos y almacenarlos en el modelo para que se realice el análisis de las emociones recogidas en los textos. También se encarga de procesar los datos del modelo para almacenarlos en un fichero cuando se comande la acción de guardar la base de datos y de la misma manera para mostrar los datos en la interfaz cuando el usuario así lo requiera.

En la clase principal, TFM_NewsFeel, se inicializa el modelo como una lista de objetos de tipo Periódico, la vista como un objeto de tipo VentanaPrincipal y el controlador como un objeto de tipo ControlPrincipal. También se realiza la primera instanciación del objeto de la clase Diccionarios.

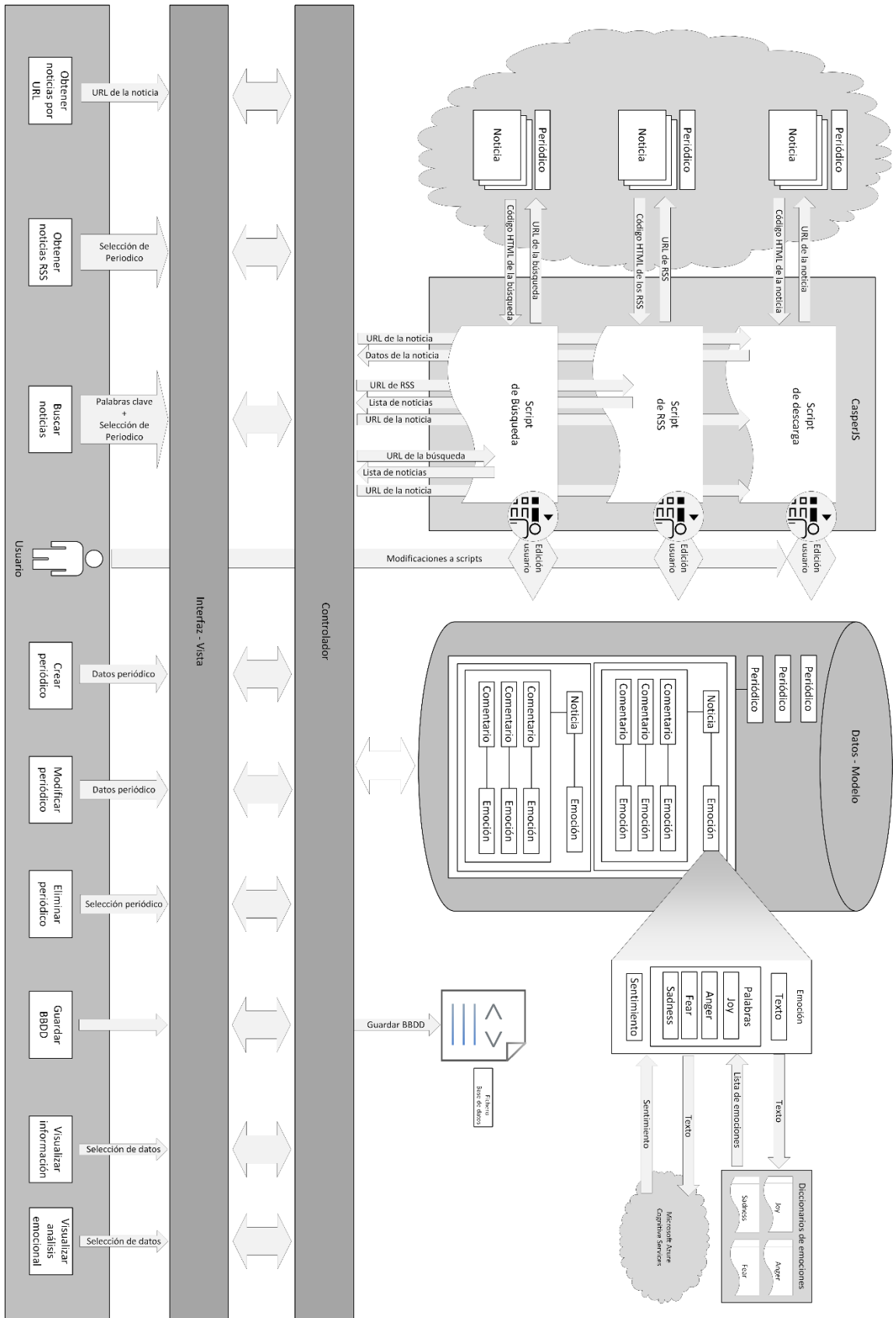


Ilustración 4. Arquitectura general del sistema

4.2 Modelo - Estructura de datos

El modelo de datos del sistema sigue un diseño basado en una estructura Periódico-Noticia-Comentario-Emoción, que permite almacenar la información de los periódicos y el análisis de emociones de una manera organizada. A esta estructura se le añaden otras clases y funcionalidades adicionales que permiten escalar el modelo de manera sencilla. A continuación, en la Ilustración 5, se muestra el diagrama de clases del modelo.

4.2.1 Diagrama de clases del Modelo

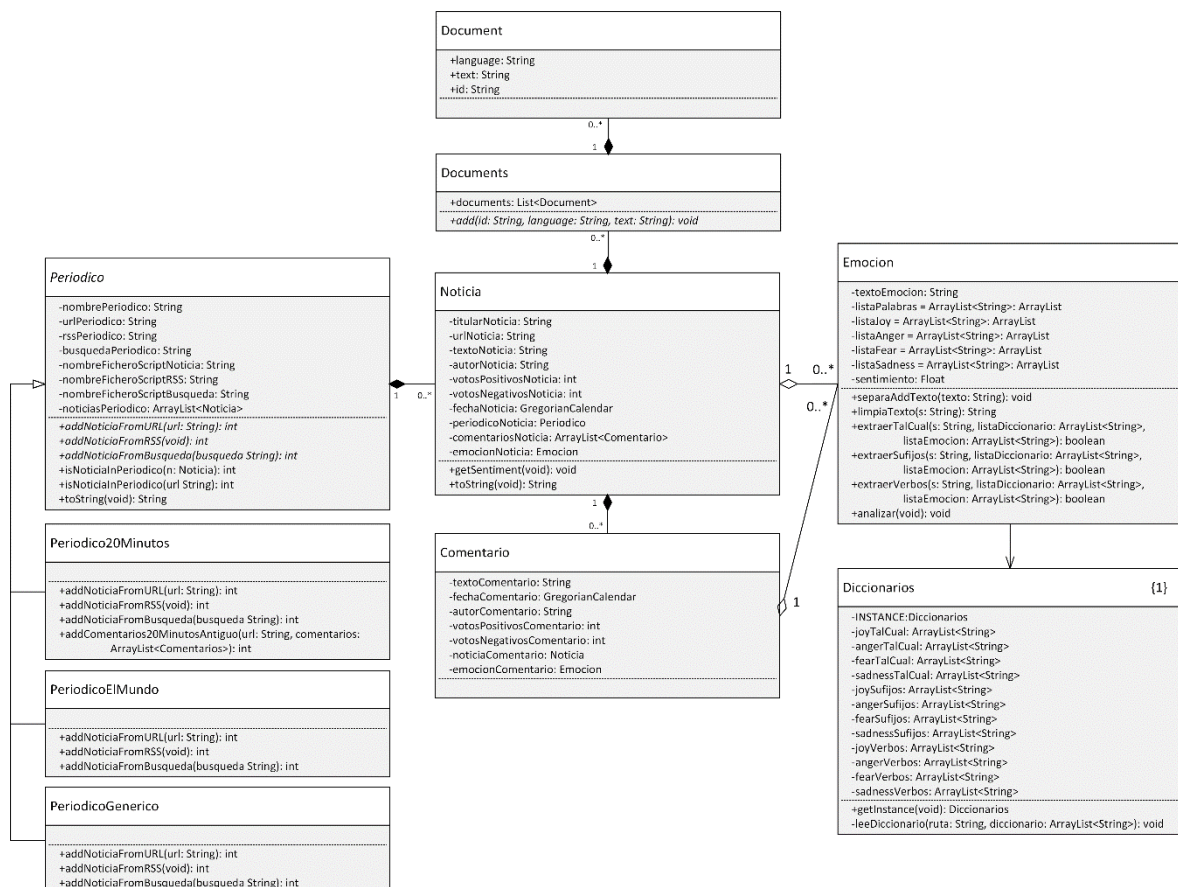


Ilustración 5. Diagrama de clases del Modelo

4.2.1.1 Clase Periodico

La clase Periodico es una clase abstracta con tres clases que heredan de ella. Los métodos abstractos de obtención de noticias son «addNoticiaFromURL», «addNoticiaFromRSS» y «addNoticiaFromBusqueda», que han de ser concretados en las clases que hereden de ésta. La clase también contiene atributos para almacenar el nombre de los scripts de descarga. Las clases Periodico20Minutos, PeriodicoElMundo y PeriodicoGenerico heredan de la clase Periodico, por lo que deben concretar los métodos abstractos en su implementación. Los atributos de la clase Periodico son los siguientes:

nombrePeriodico: Nombre del periódico.

urlPeriodico: Dirección URL base del periódico.

rssPeriodico: Dirección de los RSS del periódico.
busquedaPeriodico: URL de búsqueda del periódico a partir de la cual se pondrán palabras clave.
nombreFicheroScriptNoticia: Nombre del fichero de descarga para una noticia.
nombreFicheroScriptRSS: Nombre del fichero de descarga de noticias por RSS.
nombreFicheroScriptBusqueda: Nombre del fichero de descarga de noticias por búsqueda.
noticiasPeriodico: Lista de noticias del periódico.

4.2.1.2 Clase Noticia

La clase noticia contiene un método llamado getSentiment con el que se obtendrá el análisis del sentimiento de la noticia y de sus comentarios. Este método necesita dos clases internas, Document y Documents, que serán utilizadas para dar formato a las peticiones web para realizar el análisis del sentimiento transmitido por los textos.

Los atributos de la clase son los siguientes:

titularNoticia: Titular de la noticia.
urlNoticia: Enlace de la noticia.
textoNoticia: Cuerpo de la noticia.
autorNoticia: Autor de la noticia.
votosPositivosNoticia: Votos positivos recibidos para esa noticia.
votosNegativosNoticia: Votos negativos recibidos para esa noticia.
fechaNoticia: Fecha de publicación de esa noticia.
periodicoNoticia: Periódico al que pertenece esta noticia.
comentariosNoticia: Lista de objetos de tipo Comentario con todos los comentarios de la noticia
emocionNoticia: Contiene la información del análisis de emociones del texto de la noticia

4.2.1.3 Clase Documento

language: Idioma en el que está el texto a enviar.
text: Texto a enviar para ser analizado.
id: Identificador con el cual es posible reconocer y asociar la respuesta obtenida.

4.2.1.4 Clase Documentos

documents: Lista de objetos de tipo Document

4.2.1.5 Clase Comentario

La clase comentario no ha sufrido ninguna modificación. Estos son los atributos de la clase:

textoComentario: Cuerpo del comentario.
fechaComentario: Fecha de publicación del comentario.
autorComentario: Autor del comentario.
votosPositivosComentario: Votos positivos recibidos para ese comentario.
votosNegativosComentario: Votos negativos recibidos para ese comentario.
noticiaComentario: Noticia a la que pertenece este comentario.
emocionComentario: Contiene la información del análisis de emociones del texto del comentario.

4.2.1.6 Clase Emocion

La clase Emocion es responsable de realizar el análisis de emociones de los textos y mantiene la información del análisis de emociones del objeto que corresponda, pudiendo ser de una Noticia o de un Comentario. El atributo sentimiento almacena el indicador de polaridad del texto. Este indicador se obtiene desde la clase Noticia, al tener que realizar de manera conjunta la obtención del análisis sentimental del texto de la noticia y de todos los comentarios, no pudiéndose hacer una vez por cada texto.

textoEmocion: Texto del comentario o de la noticia a ser analizado.

listaPalabras: Lista con cada palabra del texto de la emoción

listaJoy, listaAnger, listaFear y listaSadness: Lista de palabras encontradas para cada emoción tras realizar el análisis de emociones.

sentimiento: Valor entre 0 y 1 que indica si el texto tiende a ser negativo o positivo respectivamente.

4.2.1.7 Clase Diccionarios

Esta nueva clase encapsula la funcionalidad de leer los diccionarios de emociones y almacenarlos en listas, para ser consultadas cuando se realice el análisis de emociones de un texto. Se trata de una clase «singleton», por lo que su creación se producirá sólo una vez, en la primera llamada, mientras que en las llamadas siguientes se obtendrán los datos creados previamente. La intención es poder leer los diccionarios sólo una vez, al crearse la clase. De esta manera se evitará leer los ficheros de diccionarios continuamente o manteniéndolos en memoria en varias listas. Además, cuenta con la ventaja de poder crear un objeto de Diccionarios en cualquier parte del código.

Los atributos de la clase son los siguientes:

INSTANCE: Instancia del objeto de la clase una vez creado.

joyTalCual: Lista de palabras de tipo TalCual de la emoción Joy.

angerTalCual: Lista de palabras de tipo TalCual de la emoción Anger.

fearTalCual: Lista de palabras de tipo TalCual de la emoción Fear.

sadnessTalCual: Lista de palabras de tipo TalCual de la emoción Sadness.

joySufijos: Lista de palabras de tipo Sufijos de la emoción Joy.

angerSufijos: Lista de palabras de tipo Sufijos de la emoción Anger.

fearSufijos: Lista de palabras de tipo Sufijos de la emoción Fear.

sadnessSufijos: Lista de palabras de tipo Sufijos de la emoción Sadness.

joyVerbos: Lista de palabras de tipo Verbos de la emoción Joy.

angerVerbos: Lista de palabras de tipo Verbos de la emoción Anger.

fearVerbos: Lista de palabras de tipo Verbos de la emoción Fear.

sadnessVerbos: Lista de palabras de tipo Verbos de la emoción Sadness.

4.3 Vista - Interfaz gráfica

La interfaz diseñada consiste básicamente en una ventana principal compuesta por un panel central en el que mostrar los diferentes paneles y un menú para navegar entre ellos. La Ilustración 6 muestra el diagrama de clases de este módulo.

4.3.1 Diagrama de clases de la Vista

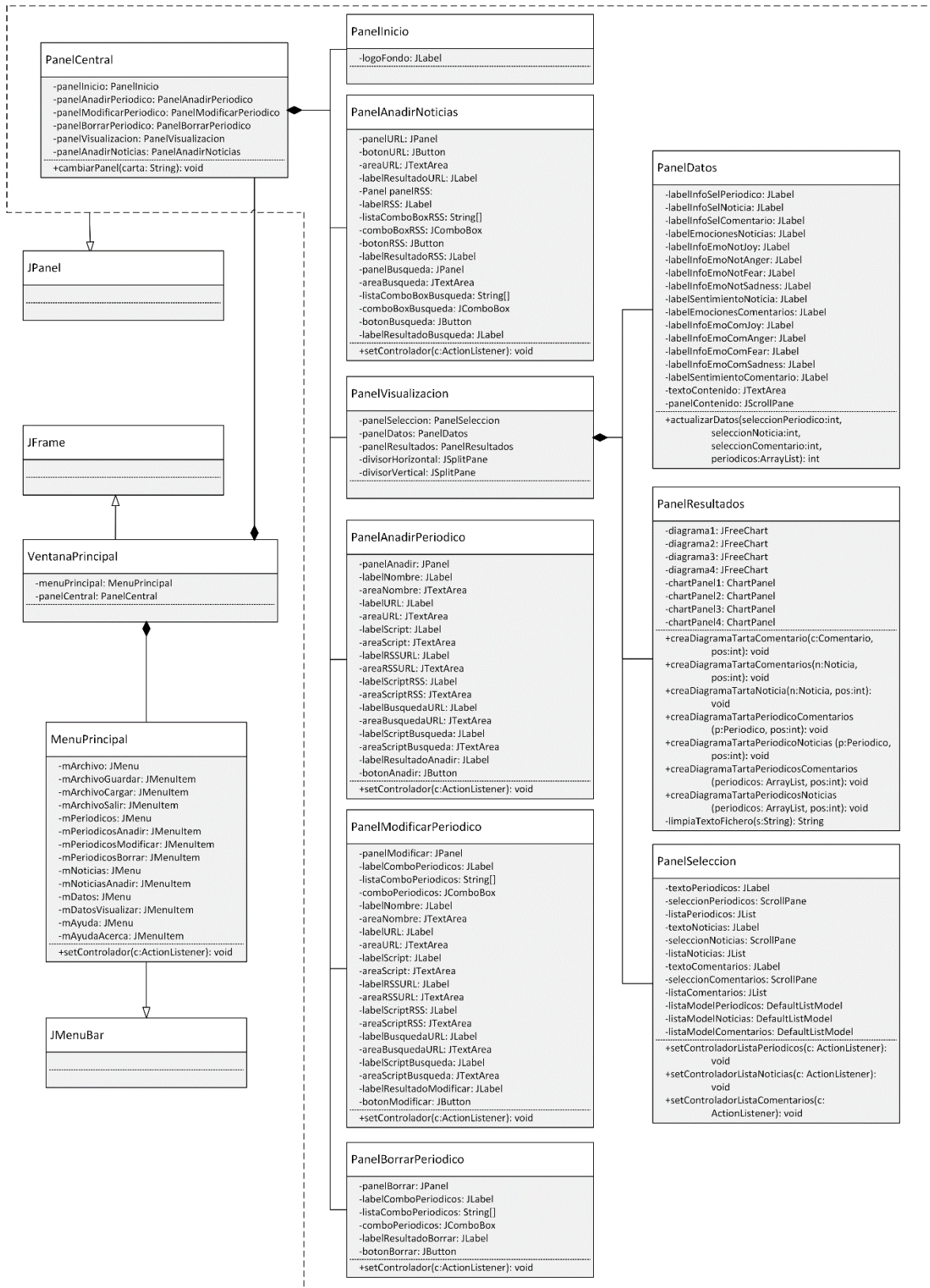


Ilustración 6. Diagrama de clases de la Vista

4.3.1.1 Clase VentanaPrincipal

Se trata de una clase que hereda de JFrame y servirá como contenedor del resto de componentes de la interfaz. Los atributos de esta clase son:

menuPrincipal: Menú diseñado para acceder a los paneles.

panelCentral: Panel que contiene el resto de los paneles de la interfaz.

4.3.1.2 Clase MenuPrincipal

Esta clase hereda de JMenuBar. Se trata de un elemento de la interfaz capaz de mostrar un menú y sus ítems. Los atributos son los siguientes:

mArchivo: Menú de archive.

mArchivoGuardar: Ítem del menú archivo para realizar la función guardar base de datos.

mArchivoCargar: Ítem del menú archivo para realizar la función cargar base de datos.

mArchivoSalir: Ítem del menú archivo para salir de la aplicación.

mPeriodicos: Menú de periódicos.

mPeriodicosAnadir: Ítem del menú periódicos para realizar la función de añadir un periódico.

mPeriodicosModificar: Ítem del menú periódicos para realizar la función de modificar un periódico.

mPeriodicosBorrar: Ítem del menú periódicos para realizar la función de borrar un periódico.

mNoticias: Menú de noticias.

mNoticiasAnadir: Ítem del menú noticias para realizar la función de añadir noticias.

mDatos: Menú de datos.

mDatosVisualizar: Ítem del menú datos para realizar la función de visualizar la información.

mAyuda: Menú de ayuda.

mAyudaAcerca: Ítem del menú ayuda para realizar la función de mostrar la información sobre la aplicación.

4.3.1.3 Clase PanelCentral

Esta clase contiene el resto de paneles de la aplicación.

panelInicio: Panel de bienvenida con el logo de la aplicación.

panelAnadirPeriodico: Panel en el que se puede añadir un periódico al sistema.

panelModificarPeriodico: Panel en el que se puede modificar la información de un periódico.

panelBorrarPeriodico: Panel en el que se puede eliminar un periódico del sistema.

panelVisualizacion: Panel en el que se puede visualizar la información de los periódicos, noticias y comentarios, así como su análisis de emociones y su representación gráfica.

panelAnadirNoticias: Panel en el que se pueden añadir noticias de diferentes maneras al sistema.

4.3.1.4 Clase *PanelInicio*

Esta clase sirve como panel de bienvenida cuando se abre la aplicación. Contiene un atributo para almacenar el logo.

logoFondo: Atributo para almacenar el logo de la aplicación

4.3.1.5 Clase *PanelAnadirPeriodico*

Desde este panel se puede introducir un periódico genérico al sistema.

panelAnadir: Panel contenedor donde añadir el resto de los elementos.
labelNombre: Etiqueta para identificar el área del nombre del periódico.
areaNombre: Área de texto para el nombre del periódico.
labelURL: Etiqueta para identificar el área de la URL base del periódico.
areaURL: Área de texto para la URL base del periódico.
labelScript: Etiqueta para identificar el área del script de descarga.
areaScript: Área de texto para el nombre del script de descarga.
labelRSSURL: Etiqueta para identificar el área de la URL de RSS.
areaRSSURL: Área de texto para la URL de RSS.
labelScriptRSS: Etiqueta para identificar el área del script de RSS.
areaScriptRSS: Área de texto para el script de RSS.
labelBusquedaURL: Etiqueta para identificar el área de búsqueda.
areaBusquedaURL: Área de texto para la URL de búsqueda.
labelScriptBusqueda: Etiqueta para identificar el área del script de búsqueda.
areaScriptBusqueda: Área de texto para el nombre del script de búsqueda.
labelResultadoAnadir: Etiqueta usada para mostrar información al usuario.
botonAnadir: Botón para añadir el periódico.

4.3.1.6 Clase *PanelModificarPeriodico*

Desde este panel se puede modificar un periódico genérico que estuviera en el sistema.

panelModificar: Panel contenedor donde añadir el resto de los elementos.
labelComboPeriodicos: Etiqueta para identificar el combo box de selección de periódico.
listaComboPeriódicos: Lista de periódicos a aparecer en el combo cox.
comboPeriodicos: Combo box para seleccionar periódico
labelNombre: Etiqueta para identificar el área del nombre del periódico.
areaNombre: Área de texto para el nombre del periódico.
labelURL: Etiqueta para identificar el área de la URL base del periódico.
areaURL: Área de texto para la URL base del periódico.
labelScript: Etiqueta para identificar el área del script de descarga.
areaScript: Área de texto para el nombre del script de descarga.
labelRSSURL: Etiqueta para identificar el área de la URL de RSS.
areaRSSURL: Área de texto para la URL de RSS.
labelScriptRSS: Etiqueta para identificar el área del script de RSS.
areaScriptRSS: Área de texto para el script de RSS.
labelBusquedaURL: Etiqueta para identificar el área de búsqueda.
areaBusquedaURL: Área de texto para la URL de búsqueda.

labelScriptBusqueda: Etiqueta para identificar el área del script de búsqueda.
areaScriptBusqueda: Área de texto para el nombre del script de búsqueda.
labelResultadoModificar: Etiqueta usada para mostrar información al usuario.
botonModificar: Botón para modificar el periódico.

4.3.1.7 Clase *PanelBorrarPeriodico*

Desde este panel se puede borrar un periódico genérico y todas sus noticias y comentarios relacionados del sistema.

panelBorrar: Panel contenedor donde añadir el resto de los elementos.
labelComboPeriodicos: Etiqueta para identificar el combo box de selección de periódico.
listaComboPeriodicos: Lista de periódicos a aparecer en el combo box.
comboPeriodicos: Combo box para seleccionar periódico
labelResultadoBorrar: Etiqueta usada para mostrar información al usuario.
botonBorrar: Botón para borrar el periódico.

4.3.1.8 Clase *PanelAnadirNoticias*

Desde este panel se pueden añadir nuevas noticias a cualquier periódico del sistema. Cuenta con tres tipos de obtención de información, por medio de URL, RSS y búsqueda.

panelURL: Panel contenedor donde situar los elementos para añadir noticias por URL.
botonURL: Botón para añadir noticias por URL.
areaURL: Área de texto para las URL de las noticias a añadir.
labelResultadoURL: Etiqueta usada para mostrar información al usuario sobre el proceso por URL
panelRSS: Panel contenedor donde situar los elementos para añadir noticias por RSS
labelRSS: Etiqueta para identificar el combo box de selección de periódico de RSS
listaComboBoxRSS: Lista de periódicos a aparecer en el combo box de RSS.
comboBoxRSS: Combo box para seleccionar periódico de RSS.
botonRSS: Botón para añadir noticias por RSS.
labelResultadoRSS: Etiqueta usada para mostrar información al usuario sobre el proceso por RSS.
panelBusqueda: Panel contenedor donde situar los elementos para añadir noticias por búsqueda.
areaBusqueda: Área de texto para las palabras clave de las noticias a buscar.
listaComboBoxBusqueda: Lista de periódicos a aparecer en el combo box de búsqueda.
comboBoxBusqueda: Combo box para seleccionar periódico de búsqueda.
botonBusqueda: Botón para añadir noticias por búsqueda.
labelResultadoBusqueda: Etiqueta usada para mostrar información al usuario sobre el proceso por búsqueda.

4.3.1.9 Clase *PanelVisualizacion*

Desde este panel se pueden ver los datos de los periódicos, noticias, comentarios y el análisis de emociones y su representación gráfica. Cuenta con tres paneles para poder seleccionar los datos, visualizar su contenido y ver la representación de los resultados del análisis de emociones.

panelSeleccion: Panel que permite seleccionar la información a visualizar.
panelDatos: Panel con la información del modelo seleccionado.
panelResultados: Panel donde mostrar diagramas con los resultados del análisis de emociones realizado sobre la información seleccionada.
divisorHorizontal: Componente que divide el panel de visualización horizontalmente, dejando arriba el panel de selección y abajo el panel de datos.
divisorVertical: Componente que divide el panel de visualización verticalmente, dejando a la derecha el panel de resultados y a la izquierda el divisor horizontal, con los paneles de selección y datos.

4.3.1.10 Clase *PanelDatos*

Desde este panel es posible ver la información de los datos seleccionados.

labelInfoSelPeriodico: Etiqueta con información de los periódicos de la selección actual.
labelInfoSelNoticia: Etiqueta con información de las noticias de la selección actual.
labelInfoSelComentario: Etiqueta con información de los comentarios de la selección actual.
labelEmocionesNoticias: Etiqueta para identificar las emociones en noticias.
labelInfoEmoNotJoy: Etiqueta con la información de palabras Joy en noticias de la selección actual.
labelInfoEmoNotAnger: Etiqueta con la información de palabras Anger en noticias de la selección actual.
labelInfoEmoNotFear: Etiqueta con la información de palabras Fear en noticias de la selección actual.
labelInfoEmoNotSadness: Etiqueta con la información de palabras Sadness en noticias de la selección actual.
labelSentimientoNoticia: Etiqueta con la información del sentimiento en noticias de la selección actual.
labelEmocionesComentarios: Etiqueta para identificar las emociones en comentarios.
labelInfoEmoComJoy: Etiqueta con la información de palabras Joy en comentarios de la selección actual.
labelInfoEmoComAnger: Etiqueta con la información de palabras Anger en comentarios de la selección actual.
labelInfoEmoComFear: Etiqueta con la información de palabras Fear en comentarios de la selección actual.
labelInfoEmoComSadness: Etiqueta con la información de palabras Sadness en comentarios de la selección actual.
labelSentimientoComentario: Etiqueta con la información del sentimiento en comentarios de la selección actual.
textoContenido: Área de texto donde se muestra la información de la noticia o comentario de la selección actual.
panelContenido: Panel contenedor donde añadir el resto de los elementos.

4.3.1.11 Clase *PanelResultados*

Desde este panel es posible ver la representación del análisis de emociones de los datos seleccionados.

Este panel se utiliza para mostrar los diagramas del análisis de emociones de los textos. En función de la selección en el panel de selección se mostrarán los resultados de unos u otros datos. Hay capacidad para hasta cuatro diagramas, siendo la posición 1 la de arriba a la

izquierda, 2 abajo a la izquierda, 3 arriba a la derecha y 4 abajo a la derecha. Los atributos de la clase son los siguientes:

diagrama1: Diagrama para el panel 1 del panel de resultados.
diagrama2: Diagrama para el panel 2 del panel de resultados.
diagrama3: Diagrama para el panel 3 del panel de resultados.
diagrama4: Diagrama para el panel 4 del panel de resultados.
chartPanel1: Panel para la posición 1 del panel de resultados.
chartPanel2: Panel para la posición 2 del panel de resultados.
chartPanel3: Panel para la posición 3 del panel de resultados.
chartPanel4: Panel para la posición 4 del panel de resultados.

4.3.1.12 Clase *PanelSeleccion*

Desde este panel es posible seleccionar el periódico, noticia o comentario para visualizar su información.

textoPeriodicos: Etiqueta para identificar la lista de selección de periódicos.
seleccionPeriodicos: Panel en el que está contenida la lista de periódicos, permitiendo realizar desplazamiento a esta lista.
listaPeriodicos: Lista de periódicos.
textoNoticias: Etiqueta para identificar la lista de selección de noticias.
seleccionNoticias: Panel en el que está contenida la lista de noticias, permitiendo realizar desplazamiento a esta lista.
listaNoticias: Lista de noticias.
textoComentarios: Etiqueta para identificar la lista de selección de comentarios.
seleccionComentarios: Panel en el que está contenida la lista de comentarios, permitiendo realizar desplazamiento a esta lista.
listaComentarios: Lista de comentarios.
listaModelPeriodicos: Modelo de datos para la lista de periódicos.
listaModelNoticias: Modelo de datos para la lista de noticias.
listaModelComentarios: Modelo de datos para la lista de comentarios.

4.4 Controlador – Lógica de aplicación

En el diseño del módulo del controlador se ha optado por una clase de controlador principal que contiene el resto de los controladores, de manera similar a como se ha realizado en la vista. Las clases de la vista en la que necesitan realizar operaciones tras ejecutar una acción contienen una clase de controlador asociada. La Ilustración 7 muestra el diagrama de clases del controlador.

En el controlador existen dos tipos en función del tipo de controlador implementado: Los que implementan la interfaz *ActionListener* y los que implementan la interfaz *ListSelectionListener*. Los primeros realizan operaciones tras detectar acciones sobre menús, botones, etc. Los segundos son referidos a las listas, en concreto, las de selección de periódico en el panel de visualización. Los controladores tienen acceso al modelo y a la vista para poder realizar las operaciones oportunas entre ellos.

4.4.1 Diagrama de clases del Controlador

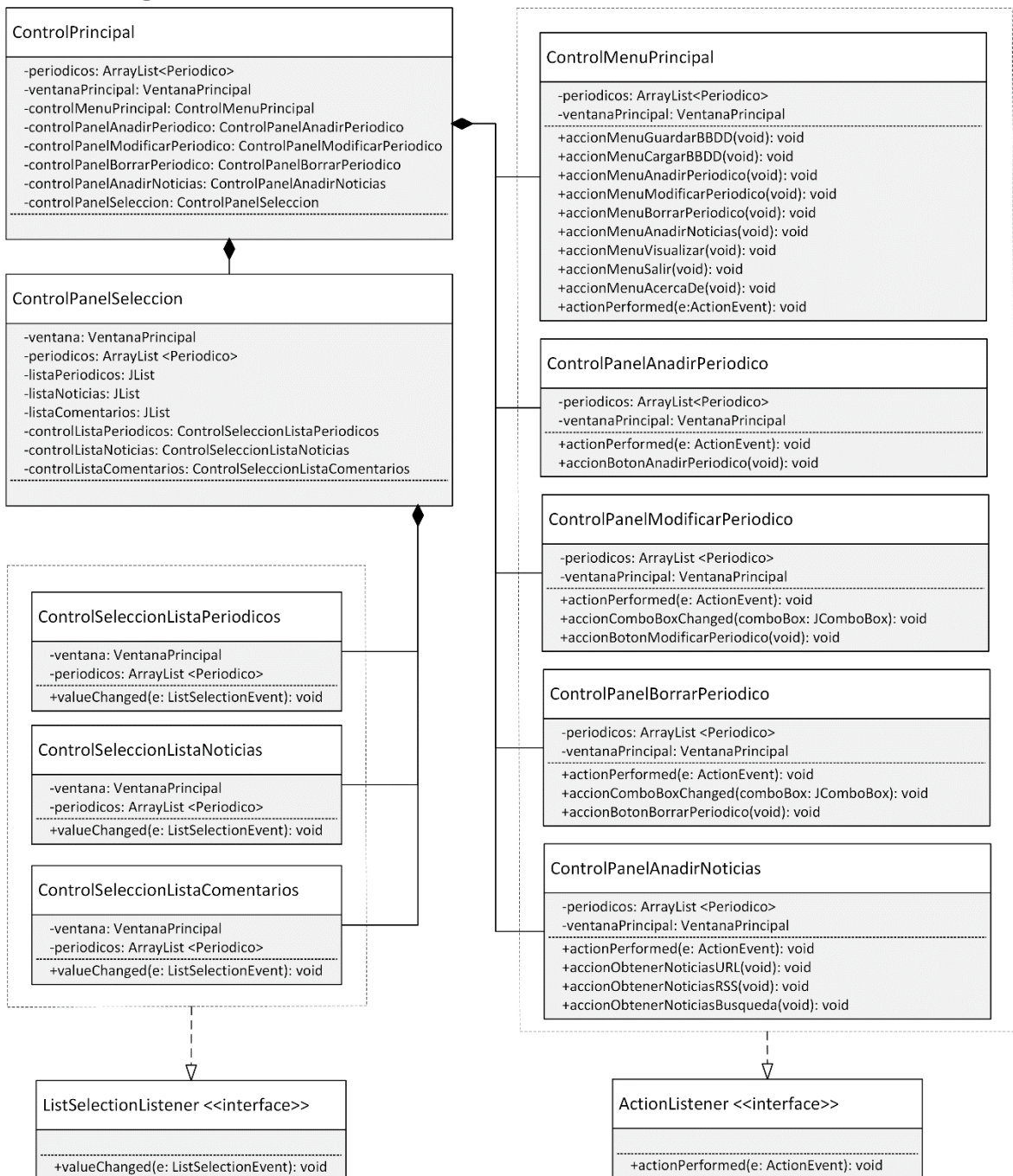


Ilustración 7. Diagrama de clases del Controlador

4.4.1.1 Clase ControlPrincipal

Esta es la clase principal del controlador. Contiene las clases controlador de los paneles de la vista necesarios. A continuación, se describen los atributos de la clase:

periodicos: Lista de periódicos. Es la clase principal del modelo de datos.

ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

controlMenuPrincipal: Contiene las acciones a realizar del menú de la aplicación.

controlPanelAnadirPeriodico: Contiene las acciones del panel de añadir un periódico.

controlPanelModificarPeriodico: Contiene las acciones del panel de modificar un periódico.
controlPanelBorrarPeriodico: Contiene las acciones del panel de borrar un periódico.
controlPanelAnadirNoticias: Contiene las acciones del panel de añadir noticias.
controlPanelSeleccion: Contiene los controladores de las listas de selección.

4.4.1.2 Clase ControlMenuPrincipal

Esta clase controla las acciones sobre los ítems del menú y realiza las operaciones oportunas.

periodicos: Lista de periódicos. Es la clase principal del modelo de datos.
ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.3 Clase ControlPanelAnadirPeriodico

Esta clase controla las acciones sobre el panel de añadir periódico y realiza las operaciones oportunas para añadir el periódico al sistema.

periodicos: Lista de periódicos. Es la clase principal del modelo de datos.
ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.4 Clase ControlPanelModificarPeriodico

Esta clase controla las acciones sobre el panel de modificar periódico y realiza las operaciones oportunas para modificar el periódico en el sistema.

periodicos: Lista de periódicos. Es la clase principal del modelo de datos.
ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.5 Clase ControlPanelBorrarPeriodico

Esta clase controla las acciones sobre el panel de borrar periódico y realiza las operaciones oportunas para borrar el periódico del sistema.

periodicos: Lista de periódicos. Es la clase principal del modelo de datos.
ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.6 Clase ControlPanelAnadirNoticias

Esta clase controla las acciones sobre el panel de añadir noticias y realiza las operaciones oportunas para añadir la noticia al sistema por el método adecuado.

periodicos: Lista de periódicos. Clase principal del modelo de datos.
ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.7 Clase ControlPanelSeleccion

Esta clase contiene las clases que controlan las listas del panel de selección. En cada una de ellas se definen operaciones que realizar cuando sus valores cambian.

periodicos: Lista de periódicos. Clase principal del modelo de datos.
ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.8 Clase ControlSeleccionListaPeriodicos

periodicos: Lista de periódicos. Clase principal del modelo de datos.

ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.9 Clase ControlSeleccionListaNoticias

periodicos: Lista de periódicos. Clase principal del modelo de datos.

ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.4.1.10 Clase ControlSeleccionListaComentarios

periodicos: Lista de periódicos. Clase principal del modelo de datos.

ventanaPrincipal: Ventana principal de la aplicación. Es la clase principal de la vista.

4.5 Periódicos

El diseño de los métodos de obtención de información plantea un reto a la hora de intentar estandarizar un procedimiento para ello.

Existen dos principales desafíos a la hora de definir un método de extracción:

- Obtener el código HTML de la noticia
- Extraer los elementos necesarios del código HTML

Se han diseñado dos clases de periódicos, cada uno siguiendo un enfoque diferente: los periódicos específicos y los periódicos genéricos. Ambos enfoques no son incompatibles entre sí, por lo que se han mantenido los dos por conveniencia para posibles proyectos futuros.

4.5.1 Periódicos específicos

Los periódicos específicos utilizan la herramienta CasperJS, usando un script externo, para navegar por la página y realizar las acciones necesarias para obtener el código HTML con todos los elementos de la noticia y sus comentarios. Este código se almacenará en un fichero que el periódico leerá y utilizará la herramienta JSoup, en las funciones de la clase, para extraer de él los elementos de las noticias y comentarios.

Este tipo de periódico puede ser útil si se sabe con seguridad que los elementos dentro del código HTML no cambiarán, ya que esta tarea esta de manera intrínseca en la definición de la clase del periódico.

Estos periódicos han de ser creados desde el inicio de la aplicación, por lo que si se quieren usar se deben crear en la clase principal de la herramienta.

4.5.2 Periódicos genéricos

Los periódicos genéricos utilizan la herramienta CasperJS, usando un script externo, para navegar por la página y realizar las acciones necesarias para obtener el código HTML, analizarlo y guardar en un fichero un dato con estructura JSON. Este fichero será leído por el periódico y almacenará directamente los datos en el modelo de datos sin realizar ningún procesamiento.

Esto permite a este tipo de periódicos no quedarse obsoletos al residir la carga del proceso de obtención del HTML y el proceso de extracción de los datos de la noticia del código en un script externo. Si uno de estos dos procesos cambiara en la parte del periódico online, cosa que ocurre a menudo, sólo habría que adaptar el script a la nueva situación pudiendo mantener la definición de la clase en el modelo de datos de la misma manera.

Gracias a ello, esto también permite realizar las siguientes tareas mientras la herramienta se encuentra en ejecución:

- Añadir periódicos
- Modificar periódicos
- Borrar periódicos

De esta manera, se puede añadir cualquier tipo de periódico a la herramienta, utilizando scripts externos para obtener los datos siempre y cuando se ajusten a los parámetros definidos por el dato en formato JSON de salida.

4.5.3 Scripts de descarga

Los scripts de descarga se ha visto que realizan distintas funciones dependiendo si están diseñados para un periódico específico o para un periódico genérico. Mientras que los parámetros de entrada al script de CasperJS son iguales, los datos obtenidos por el script, usados como salida, varían.

4.5.3.1 Script para periódicos específicos

Script de descarga:

casperjs.exe «nombreScript» «URL» «nombreFicheroSalida»

En estos periódicos el nombre del fichero de salida está preestablecido.

Fichero de salida Periodico20Minutos: «noticia_20minutos_temp.html»

Fichero de salida PeriodicoElMundo: «noticia_ElMundo_temp.html»

El fichero de salida debe contener el código HTML con los datos de la noticia y todos los comentarios con los datos en crudo.

Script de RSS:

casperjs.exe «nombreScript» «URLRSS» «nombreFicheroSalida»

En estos periódicos el nombre del fichero de salida está preestablecido.

Fichero de salida Periodico20Minutos : «rss_20minutos_temp.html»

Fichero de salida PeriodicoElMundo: «rss_ElMundo_temp.html »

El fichero de salida debe contener el código HTML de la página de RSS.

Script de búsqueda:

casperjs.exe «nombreScript» «URLBusqueda_conParametros» «nombreFicheroSalida»

En estos periódicos el nombre del fichero de salida está preestablecido.

Fichero de salida Periodico20Minutos : «busqueda_20minutos_temp.html»

Fichero de salida PeriodicoElMundo: « busqueda_elmundo_temp.html»

El fichero de salida debe contener el código HTML de la página de búsqueda realizada.

4.5.3.2 Script para periódicos genéricos

Script de descarga:

casperjs.exe «nombreScript» «URL» temp_noticia_generica.json

El fichero de salida debe contener los datos de la noticia y sus comentarios con el siguiente formato:

```
{
  "titularNoticia" : "Titular de la noticia",
  "textoNoticia" : "Texto de la noticia",
  "autorNoticia" : "Autor de la noticia",
  "votosPositivosNoticia" : votos (numero entero),
  "votosNegativosNoticia" : votos (numero entero),
  "fechaNoticia" : "HH:MM DD/MM/AAAA",
  "comentariosNoticia" :
  [
    {
      "textoComentario" : "Texto del comentario",
      "fechaComentario" : "HH:MM DD/MM/AAAA",
      "autorComentario" : "Autor del comentario",
      "votosPositivosComentario" : votos (numero entero),
      "votosNegativosComentario" : votos (numero entero)
    },
    {
      "textoComentario" : "Texto del comentario",
      "fechaComentario" : "HH:MM DD/MM/AAAA",
      "autorComentario" : "Autor del comentario",
      "votosPositivosComentario" : votos (numero entero),
      "votosNegativosComentario" : votos (numero entero)
    }
  ]
}
```

Ilustración 8. Formato JSON de salida para script de descarga

Nota: La fecha debe tener el formato 24h. Los votos deben tener un formato numérico entero.

Script de RSS:

casperjs.exe «nombreScript» «URLRSS» «temp_rss_noticia_generica.json»

El fichero de salida debe contener los datos de los enlaces de RSS con el siguiente formato:

```
{
  "enlacesBusqueda" :
  [
    {
      "enlace" : "URL"
    },
    {
      "enlace" : "URL"
    },
    {
      "enlace" : "URL"
    }
  ]
}
```

Ilustración 9. Formato JSON de salida para script de RSS

Script de búsqueda:

casperjs.exe «nombreScript» «URLBusqueda_conParametros»
temp_busqueda_noticia_generica.json

El fichero de salida debe contener los datos de los enlaces de noticias encontradas con el siguiente formato:

```
{
  "enlacesRSS" :
  [
    {
      "enlace" : "URL"
    },
    {
      "enlace" : "URL"
    },
    {
      "enlace" : "URL"
    }
  ]
}
```

Ilustración 10. Formato JSON para script de búsqueda

4.6 Obtención de noticias y análisis de emociones

4.6.1 Obtención de noticias

Para obtener las noticias y todos sus comentarios se ha establecido tres métodos de obtención:

- **Obtención de noticias por URL, función addNoticiasFromURL:**
Permite descargar una noticia utilizando para ello la dirección o URL de la noticia en el periódico online. No es necesario especificar el periódico al que pertenece, puesto que se detectará utilizando la URL de la noticia. Los datos se extraerán utilizando el script de descarga. Los datos obtenidos de la noticia y comentarios se almacenarán en el modelo de datos y posteriormente se realizará el análisis de emociones de la noticia y los comentarios.

- **Obtención de noticias por RSS, función `addNoticiasFromRSS`:**
Permite descargar varias noticias que se encuentren en la lista de RSS proporcionada por el periódico online. Para ello primero se obtendrán los enlaces de las noticias en los RSS utilizando el script y después se descargará cada noticia por el método de URL. Se puede seleccionar obtener las noticias de RSS de un periódico determinado o bien si se quiere descargar de todos los periódicos.
- **Obtención de noticias por búsqueda, función `addNoticiasFromBusqueda`:**
Permite descargar varias noticias que resulten de una búsqueda mediante unas palabras clave. Para ello primero se obtendrán los enlaces de las noticias encontradas utilizando el script y después se descargará cada noticia utilizando la URL por el método de URL. Al igual que en los RSS, se puede seleccionar obtener los enlaces de un periódico determinado o de todos los periódicos.

4.6.2 Análisis de emociones

El análisis de emociones consta de dos fases. La primera es el análisis de emociones encontradas en los textos de las noticias y periódicos. La segunda es encontrar el sentimiento o polaridad de cada uno de esos textos, para saber si es un comentario positivo, negativo o neutro.

4.6.2.1 Análisis de emociones

Una vez que se tienen los textos de la noticia y de los comentarios descargados se realizará el análisis de emociones.

El análisis consiste en comparar cada palabra del texto con las palabras en unos diccionarios de emociones, con la finalidad de detectar las emociones en el texto.

Existen tres tipos de diccionarios en función de las emociones que representan sus palabras. Joy (alegría), Anger (Ira), Fear (Miedo) y Sadness (Tristeza). Si coincide la palabra del texto con la palabra del diccionario de una emoción, la palabra pasará a formar parte de una lista de palabras detectadas para esa emoción.

Para dar soporte a la lectura de los diccionarios se ha creado la clase `Diccionarios`. Se trata de una clase singleton, que tiene la capacidad de una vez creada, poderse instanciar varias veces. Esto se ha realizado así para evitar leer constantemente de los ficheros de diccionarios, manteniendo todos los datos en un objeto definido.

Existen varios tipos de diccionarios, en función de cómo realizar la comparación, para cada emoción: palabras `TalCual`, palabras `Sufijos`, y palabras `Verbos`.

Las palabras `TalCual` son palabras que se compararán sin realizar ninguna modificación en ellas, más allá de añadir la forma en plural.

Las palabras `Sufijos` son palabras que podrían tener muchas formas variadas en función de su terminación. Se une la raíz de las palabras con un conjunto de sufijos. De esta manera, cada palabra en este tipo de diccionario se multiplica.

Las palabras `Verbos` son palabras que, como su nombre indica, se trata de verbos. Estos verbos vienen en su forma de infinitivo. Para poder detectar todas las formas verbales de

estas palabras en los textos se compondrán todas las formas verbales de ellas a partir de tres conjuntos de terminaciones, en función de la conjugación del verbo del diccionario.

Tras realizar el análisis, el resultado serán listas de palabras detectadas por cada emoción. Estas listas se guardarán en el modelo de datos en los objetos Emoción para cada noticia y cada comentario.

4.6.2.2 Análisis del sentimiento

El sentimiento de un texto indica cuánto de positivo o negativo es ese texto. Esto se representa mediante un valor entre 1 y 0 respectivamente. Este análisis forma parte de la API «Text Analytics» y se realiza en los servidores de «Microsoft Cognitive Services». Los textos son analizados con el uso de algoritmos de aprendizaje automático e inteligencia artificial directamente en la nube.

La manera ideal de haber realizado estos análisis hubiera sido obteniendo el sentimiento de cada texto a medida que se vayan obteniendo del medio online. Sin embargo, debido a las limitaciones de uso de la versión gratuita de Text Analytics se ha realizado de una manera ligeramente diferente. Estas restricciones de uso, entre otras, conlleva que haya un límite de análisis de 5.000 peticiones al mes, por lo que hacer un análisis de cada texto teniendo en cuenta que sólo una noticia puede tener cientos de comentarios, se superaría muy fácilmente el límite. No obstante, se pueden agrupar varios textos para ser incluidos en una misma petición, evitando así superar el límite.

Por lo tanto, para realizar el análisis de los textos se enviarán al servidor el texto de la noticia junto al texto de todos los comentarios.

El formato de envío de información debe ser el siguiente:

```
{
  "documents":
  [
    {
      "language": "es",
      "id": "1",
      "text": "Esto es un texto a analizar."
    },
    {
      "language": "es",
      "id": "2",
      "text": "Esto es otro texto a analizar."
    }
  ]
}
```

Ilustración 11. Formato JSON de envío de textos para análisis del sentimiento

Consiste en una variable «documents» que contiene una lista de un conjunto de variables: «language», «text» y «id». La primera variable «language» debe tener el valor del idioma en el que está escrito el texto. En este caso el valor para el castellano es “es”. La variable «text» debe contener el texto a analizar. Por último, la variable «id» contiene un identificador con el cual se pueden asociar las respuestas.

Para dar este formato a los textos de las noticias y comentarios se han utilizado dos clases de los ejemplos de código iniciales de la API de Text Analytics, Document y Documents. Estas clases se han declarado de manera interna en la clase Noticia. Una vez se haya creado el objeto de tipo Noticia con todos los textos se poblarán los objetos Documents y Document para generar el documento JSON a enviar al servidor.

Una vez que el texto se ha analizado se obtendrá la respuesta, que tendrá el siguiente formato:

```
{
  "documents": [{
    "score": 0.51099812984466553,
    "id": "1"
  }, {
    "score": 0.5620684027671814,
    "id": "2"
  }],
  "errors": []
}
```

Ilustración 12. Formato JSON de respuesta del análisis del sentimiento de textos

El texto de respuesta consiste en otra variable «documents» con una lista de otro conjunto de variables: «score» y «id». La variable «score» nos dice el valor del análisis entre 1 y 0. Con la variable «id» se pueden asociar las respuestas a los textos enviados. Por último, hay un elemento «errors» en el que saldrá la información de errores si se produce alguno.

Por último, se procesa el texto de respuesta con los resultados, asignando al objeto Emocion el valor del sentimiento del texto. Esto se realiza para el objeto Emocion de la Noticia y todos los objetos Emocion de los objetos Comentario de la Noticia.

4.7 Interfaz gráfica de usuario

4.7.1 Ventana principal y menús

La ventana principal contiene el y un panel central

La interfaz gráfica está diseñada basada en una ventana de aplicación. Esta ventana contendrá un menú desde donde se accederá a los distintos paneles y un panel central en el que se mostrarán los diferentes paneles de la aplicación.

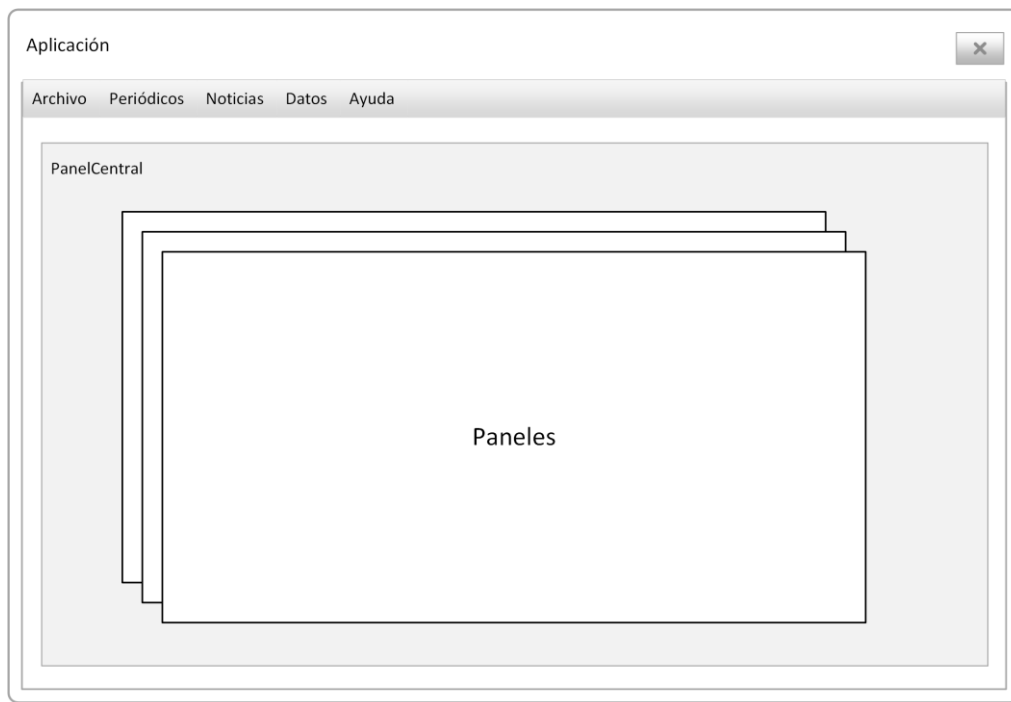


Ilustración 13. Ventana principal de la interfaz gráfica de usuario

4.7.2 Gestión de periódicos

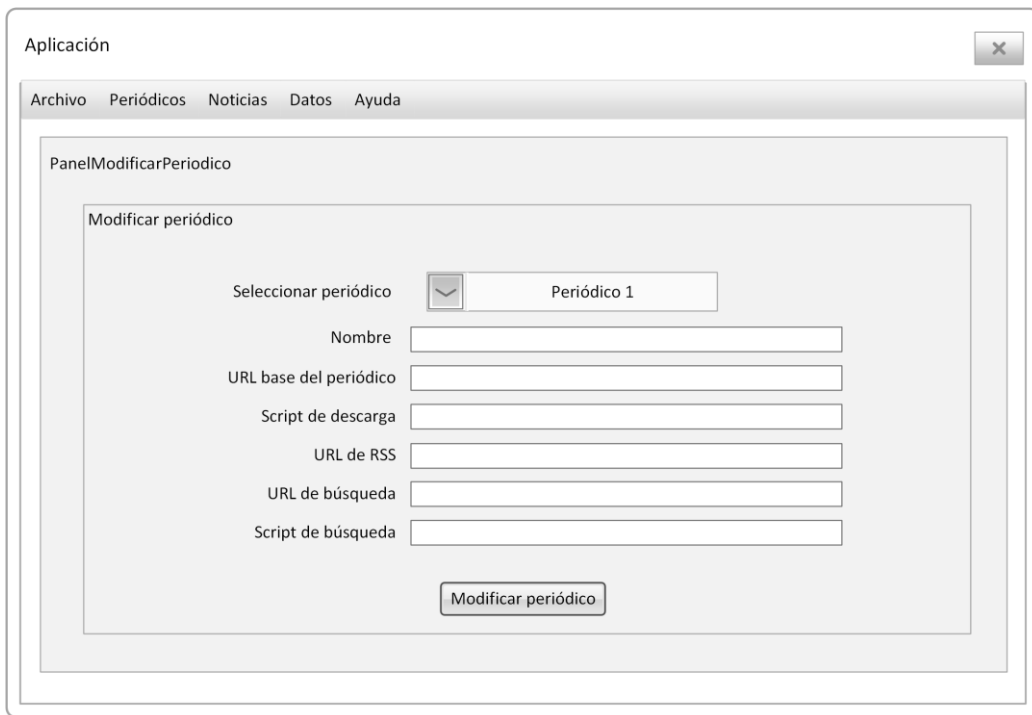
A través del menú se puede acceder a los paneles de gestión de periódicos.

4.7.2.1 Añadir un periódico

The image shows the same 'Aplicación' window, but the 'Periódicos' menu item is selected. The main content area now displays 'PanelAñadirPeriodico'. Inside this panel, there is a sub-section titled 'Añadir periódico'. This section contains six text input fields arranged vertically, each with a label to its left: 'Nombre', 'URL base del periódico', 'Script de descarga', 'URL de RSS', 'URL de búsqueda', and 'Script de búsqueda'. Below these input fields is a single button labeled 'Añadir periódico'.

Ilustración 14. Panel para añadir un periódico

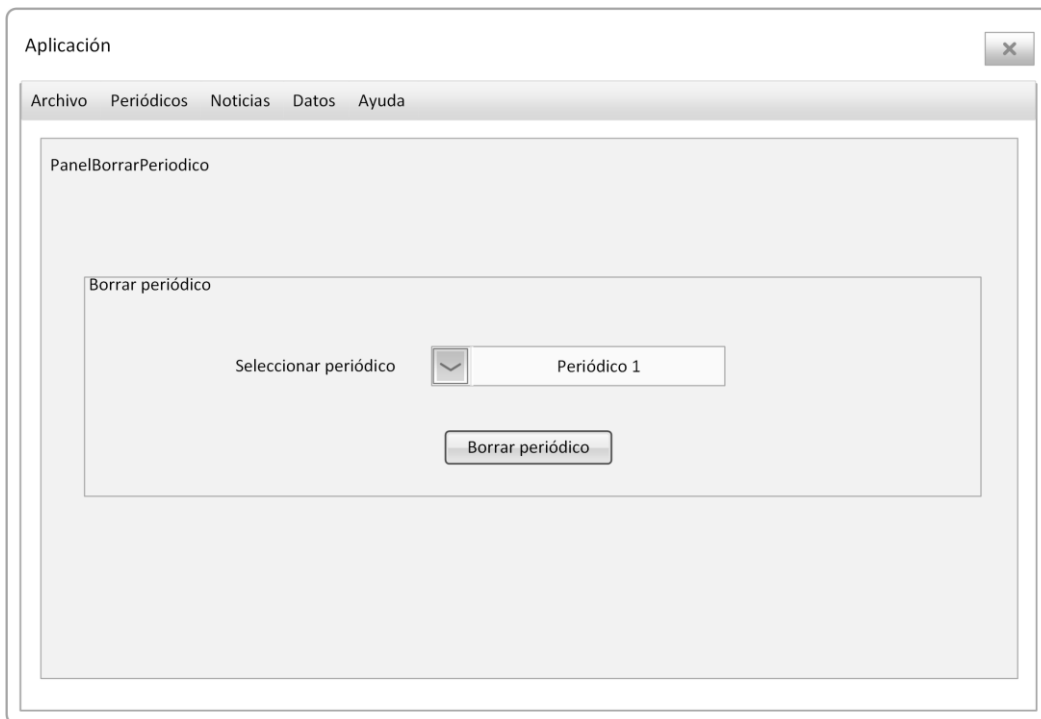
4.7.2.2 Modificar un periódico



The screenshot shows a window titled 'Aplicación' with a menu bar containing 'Archivo', 'Periódicos', 'Noticias', 'Datos', and 'Ayuda'. The main content area is titled 'PanelModificarPeriodico'. Inside this panel, there is a sub-section titled 'Modificar periódico'. This section contains a dropdown menu labeled 'Seleccionar periódico' with a downward arrow and the text 'Periódico 1'. Below this are six text input fields labeled 'Nombre', 'URL base del periódico', 'Script de descarga', 'URL de RSS', 'URL de búsqueda', and 'Script de búsqueda'. At the bottom of the sub-section is a button labeled 'Modificar periódico'.

Ilustración 15. Panel para modificar un periódico

4.7.2.3 Borrar un periódico



The screenshot shows a window titled 'Aplicación' with a menu bar containing 'Archivo', 'Periódicos', 'Noticias', 'Datos', and 'Ayuda'. The main content area is titled 'PanelBorrarPeriodico'. Inside this panel, there is a sub-section titled 'Borrar periódico'. This section contains a dropdown menu labeled 'Seleccionar periódico' with a downward arrow and the text 'Periódico 1'. Below this is a button labeled 'Borrar periódico'.

Ilustración 16. Panel para borrar un periódico

4.7.3 Añadir noticias

Desde este panel se pueden añadir las noticias de distintos modos.

The screenshot shows a window titled 'Aplicación' with a menu bar containing 'Archivo', 'Periódicos', 'Noticias', 'Datos', and 'Ayuda'. The main content area is titled 'PanelAñadirNoticias' and contains three sections for adding news:

- Añadir noticias a partir de URLs:** A text input field and a button labeled 'Obtener noticias desde URL'.
- Añadir noticias a partir de RSS:** A dropdown menu labeled 'Seleccionar periódico' with 'Periódico 1' selected, and a button labeled 'Obtener noticias desde RSS'.
- Añadir noticias a partir de palabras clave:** A text input field, a dropdown menu labeled 'Periódico 1', and a button labeled 'Buscar noticias'.

Ilustración 17. Panel para añadir noticias

4.7.4 Visualización de datos

En el panel de visualización de datos hay tres paneles. Panel de selección, panel de datos y panel de visualización.

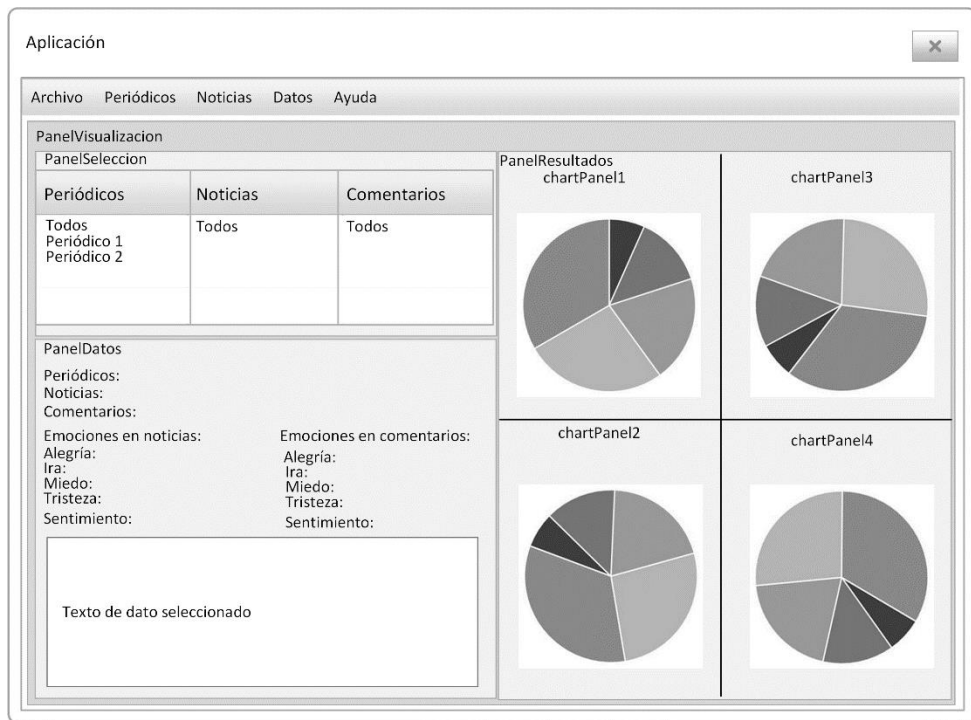


Ilustración 18. Panel para visualizar datos y resultados

5 Desarrollo

Este proyecto ha sido desarrollado en el lenguaje de programación Java de acuerdo a las especificaciones en la sección 3.8 Requisitos, y al diseño realizado en la sección 4 Diseño.

5.1 Proyecto de Java

El proyecto se ha estructurado separando los archivos fuentes de la clase principal, de vista modelo y controlador en diferentes paquetes: La clase principal se ha ubicado en el paquete `tfm_newsfeel`. Los controladores en el paquete `tfm_newsfeel.controllers`, el modelo en `tfm_newsfeel.models` y la vista en `tfm_newsfeel.views`.

5.2 Clase principal

La aplicación está desarrollada sobre una clase principal, llamada `TFM_NewsFeel.java`. En el método `main` de esta clase se encuentra una llamada al método `SwingUtilities.invokeLater`. Este método permite crear un hilo de ejecución de Swing para poder cargar la interfaz de usuario. En él se debe sobrecargar el método `run()` y es aquí donde podremos empezar con el código de la aplicación.

Nada más ejecutar la aplicación lo primero es crear la variable donde se guardará el modelo de datos, una lista de objetos Periódico. A continuación, se intentará leer la base de datos por defecto, que se encontrará en el directorio raíz del proyecto con el nombre «`noticias.xml`». Si existe el fichero se lee y el contenido es asignado a la variable de la lista de periódicos, donde la poblará con todos los objetos que tuviera. Si el fichero no existe se crearán los periódicos `Periodico20Minutos` y `PeriodicoElMundo` y se guardarán los objetos de la lista de periódicos creando el fichero «`noticias.xml`». Estos periódicos son de tipo específicos y sólo se pueden crear en este punto de ejecución. Más adelante se verá como añadir periódicos genéricos en tiempo de ejecución.

Una vez se tenga el modelo de datos con los datos cargados o los periódicos iniciales se instanciará la clase singleton `Diccionarios`, para leer los diccionarios y guardarlos en el objeto.

A continuación, se creará el objeto de la clase `VentanaPrincipal`. Esta clase es la que contiene todos los paneles de la vista.

Por último, se creará el objeto de la clase `ControlPrincipal`, conteniendo todos los controladores. La clase principal tiene como parámetros el modelo y la vista, para poder acceder a ellos cuando sea necesario.

Por último, cuando se ejecute el hilo de Swing aparecerá la ventana de la interfaz gráfica y el usuario podrá comenzar a realizar acciones en ella.

5.3 Modelo

El modelo de datos se ha creado en una lista de objetos periódicos en la clase principal del proyecto.

Al crear un periódico sólo se crea su objeto. La lista de noticias que contiene está vacía aún. Cuando se añade una noticia por cualquier método de los disponibles la noticia contiene también el objeto Emocion y los comentarios si la noticia los tuviera, junto sus correspondientes objetos Emocion.

5.4 Vista

Cuando se crea el objeto de la clase VentanaPrincipal en su constructor comienza a crear los distintos objetos de los elementos que contiene y cada uno de éstos crea también los que contienen ellos. Después de la propagación de creación de los objetos contenidos se puede acceder a cualquier elemento de la vista.

5.4.1 El panel central

El panel central es un elemento clave de la interfaz. Contiene todos los paneles que se mostrarán en la ventana de la aplicación. Este panel tiene un diseño especial llamado CardLayout. Con este diseño se pueden alternar la visualización de los paneles como si fueran cartas apiladas. Cada vez que desde el menú en la interfaz se haga click para cambiar a otro panel se llama a la función cambiarPanel de esta clase. El parámetro de entrada es un String identificativo con el que se asocia cada panel. De este modo se muestra el panel elegido.

5.4.2 El panel visualización

Otro elemento clave de la interfaz es el panel de visualización. Dentro de este panel se encuentran tres paneles más de la vista: el panel de selección, el panel de datos y el panel de resultados.

5.4.2.1 El panel selección

En el panel de selección se encuentran una lista de periódicos, otra lista de noticias y otra lista de comentarios. En función de la selección de esas listas se mostrará la información del panel de datos y los diagramas del panel de resultados.

Cada lista tiene siempre un elemento inicial, TODOS, que representa a seleccionar todos los elementos de esa lista. Cuando se selecciona un elemento de la lista automáticamente en la lista siguiente se selecciona el elemento TODOS. Por ejemplo, si hubiera una noticia seleccionada y en la lista de periódicos se selecciona otro periódico no tendría sentido que se mantuviera la noticia del otro periódico seleccionada ya que además no pertenece a ese periódico.

Otra peculiaridad del comportamiento de este panel es que al seleccionar el elemento TODOS de una lista en la siguiente lista sólo aparece el elemento TODOS. Esto es para no mezclar noticias de distintos periódicos en la lista y comentarios de distintas noticias.

Esto se ha implementado en las clases del controlador de listas ControlSeleccionListaPeriodicos, ControlSeleccionListaNoticias, ControlSeleccionListaComentarios, en la función valueChanged.

Cada vez que se selecciona un elemento, al modificar las listas siguientes, siempre se acaba modificando la lista de comentarios. En la función `valueChanged` de la clase `ControlSeleccionListaComentarios` se actualizan los datos del panel de datos y se crean los diagramas adecuados para el periódico, noticia y comentario seleccionados.

5.4.2.2 El panel de resultados

En este panel se añaden los diagramas del análisis de emociones en función del periódico, noticia y comentario seleccionado en el panel de selección.

Hay cuatro posibles escenarios:

Selección de todos los periódicos, todas las noticias y todos los comentarios: Se llamarán a las funciones `creaDiagramaTartaPeriodicosNoticias` y `creaDiagramaTartaPeriodicosComentarios`, que crean un diagrama con todas las emociones encontradas en las noticias de todos los periódicos y todas las emociones encontradas en los comentarios de todos los periódicos respectivamente.

Selección de un periódico, todas las noticias y todos los comentarios: Se llamarán a las funciones `creaDiagramaTartaPeriodicoNoticias` y `creaDiagramaTartaPeriodicoComentarios`, que crean un diagrama con todas las emociones encontradas en las noticias del periódico seleccionado y todas las emociones encontradas en los comentarios del periódico seleccionado respectivamente.

Selección de un periódico, una noticia y todos los comentarios: Se llamarán a las funciones `creaDiagramaTartaNoticia` y `creaDiagramaTartaComentarios`, que crean un diagrama con todas las emociones encontradas en una noticia de un periódico y todas las emociones encontradas en los comentarios de una noticia de un periódico respectivamente.

Selección de un periódico, una noticia y un comentario: Se llamará a la función `creaDiagramaTartaComentario`, que crea un diagrama con todas las emociones encontradas en el comentario de una noticia de un periódico.

Además, este panel tiene cuatro `ChartPanel`. Las funciones anteriores tienen un parámetro, posición, en el que se le puede indicar el `ChartPanel` en el que pueden ir como se muestra en la Ilustración 18. Panel para visualizar datos y resultados.

5.5 Controlador

Cuando se crea el objeto de la clase `ControlPrincipal`, al igual que la clase `Ventana principal`, en su constructor comienza a crear los distintos objetos de los elementos que contiene y cada uno de éstos crea también los que contienen ellos. Después de la propagación de creación de los objetos contenidos se puede acceder a cualquier elemento del controlador.

6 Ejemplos

6.1 Descarga de noticias desde una búsqueda

Para descargar noticias de todos los periódicos utilizando una búsqueda de unas palabras clave se puede realizar desde el panel «Añadir noticias».

Para realizar la descarga, introducir las palabras clave a buscar en el cuadro de la parte inferior del panel, y seleccionar el periódico de donde obtenerlas. En el ejemplo mostrado en la Ilustración 19, se desea buscar noticias relacionadas con la palabra clave «Urdangarin» y se han seleccionado todos los periódicos.

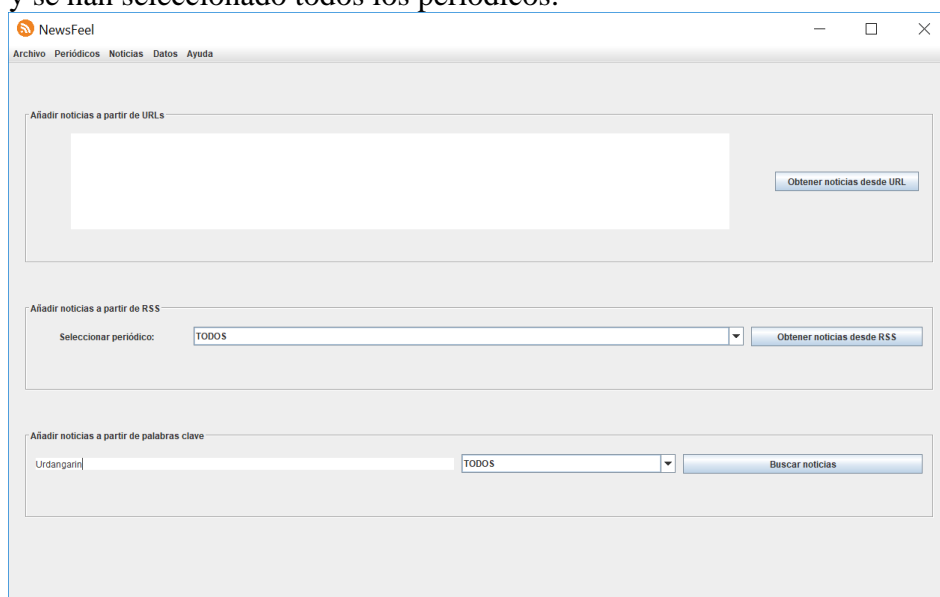


Ilustración 19. Panel para añadir noticias mediante búsqueda

Una vez se hayan obtenido las noticias aparecerá un mensaje confirmando la descarga y el número de noticias obtenidas.

Por último, cuando se vuelva al panel de visualización aparecerán todas las noticias sobre esa búsqueda en cada periódico. Como se puede ver en la Ilustración 20, en la parte izquierda aparece la lista de los periódicos disponibles en la herramienta. En la lista de en medio aparecen las noticias encontradas en el ejemplo que nos ocupa (la búsqueda de noticias relacionadas con Urdangarin). Debajo se encuentra información numérica sobre los datos seleccionados; información del periódico, noticia y comentario seleccionado, información de las emociones encontradas en textos de noticias y las emociones encontradas en textos de comentarios. En la parte derecha de la ilustración se muestran gráficamente los resultados del análisis emocional de la selección de la izquierda: se ha encontrado un 100% de anger en la noticia. Con respecto a los comentarios, analizados en conjunto se obtiene que un 47% refleja joy, otro 27% refleja fear, otro 13% sadness y por último otro 13% anger..

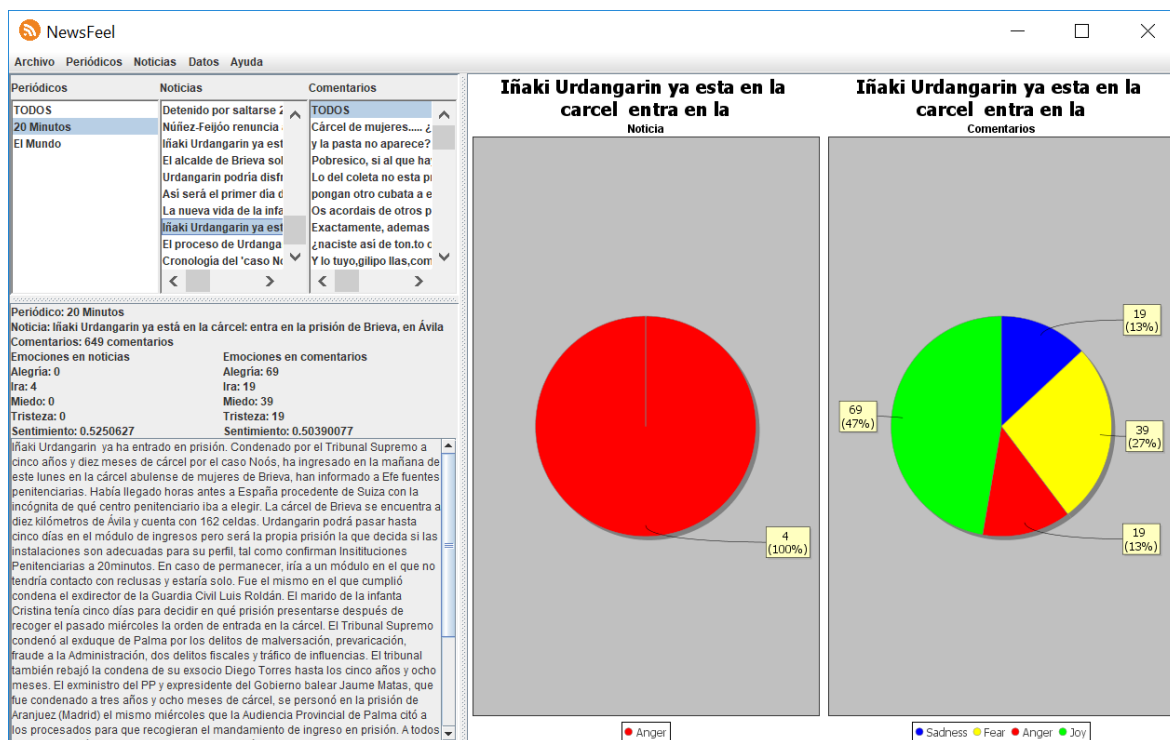


Ilustración 20. Datos de noticia del periódico online 20 minutos

6.2 Comparación de noticias similares en periódicos distintos

En este apartado se van a comparar dos noticias similares obtenidas desde distintos periódicos. Se trata de dos noticias obtenidas mediante la búsqueda del ejemplo anterior.

6.2.1 Noticia en el periódico 20 Minutos

En la Ilustración 20 se pueden ver los datos de la noticia de 20 Minutos. Entre los sentimientos encontrados en el texto de la noticia sólo se encuentran sentimientos de ira. Por otra parte, los sentimientos encontrados en los comentarios casi el 50% son de alegría.

En uno de los comentarios de esta noticia se han encontrado un sentimiento de alegría y otro de ira, , ver Ilustración 21. Además, el valor del sentimiento indica que el texto del comentario tiene un tono positivo. Leyendo el texto del comentario se puede corroborar que los valores obtenidos mediante el análisis de sentimiento son acertados en este caso.

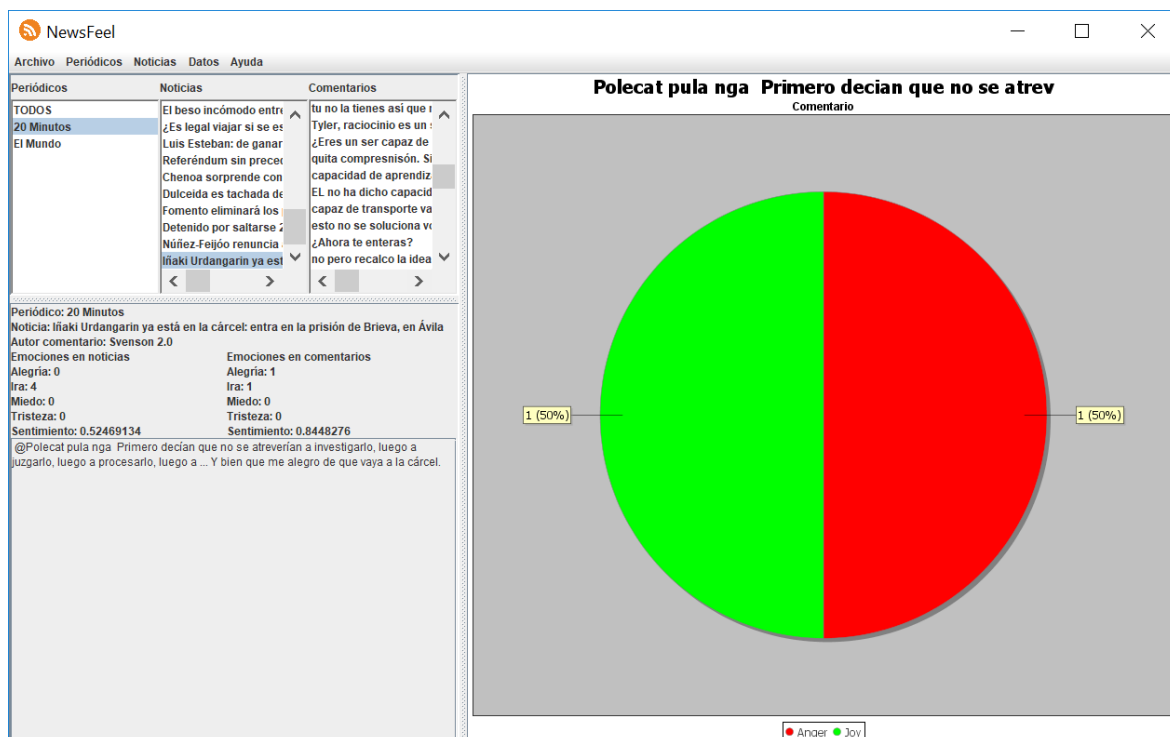


Ilustración 21. Datos de comentario del periódico 20 Minutos

6.2.2 Noticia en el periódico El Mundo

En la Ilustración 22 se pueden ver los datos de la noticia del periódico El Mundo. Entre los sentimientos encontrados en el texto de la noticia se encuentran sentimientos de ira y miedo. En cambio, la mayoría de los sentimientos encontrados en los comentarios son de miedo y en segundo lugar de alegría. En este caso también se observa un texto con sentimiento neutral, y la media de sentimientos de los textos correspondientes a los comentarios también se acerca al punto medio.

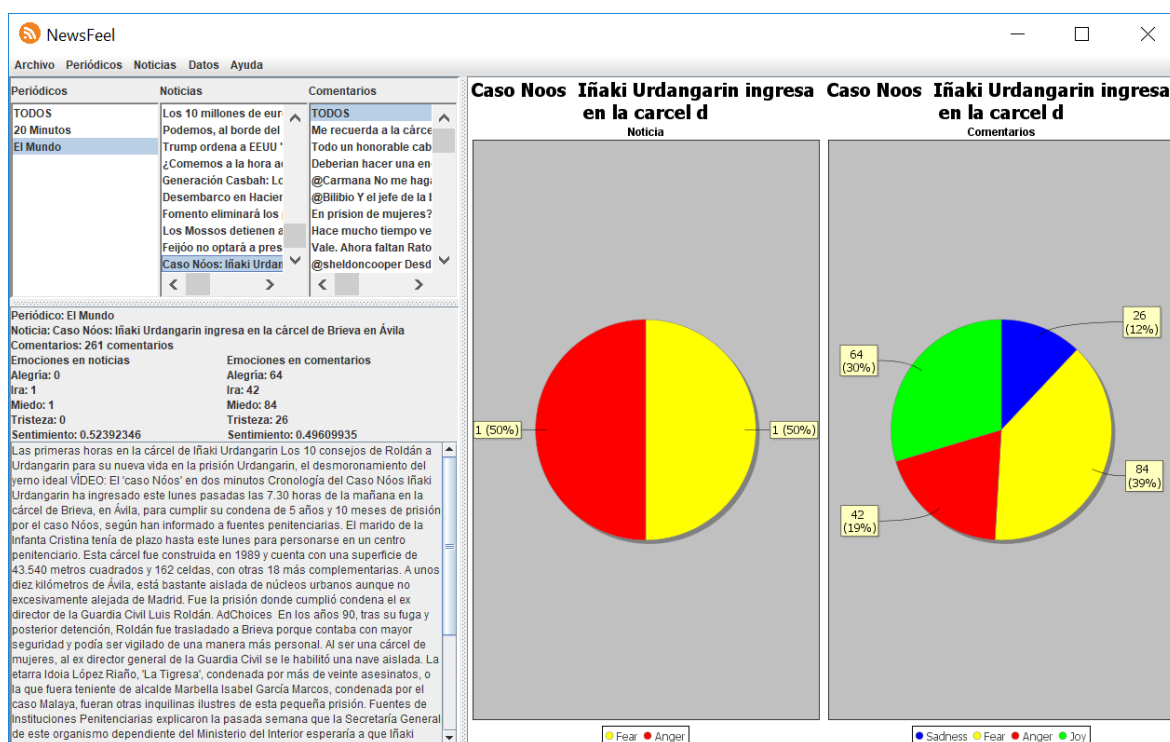


Ilustración 22. Datos de noticia del periódico online El Mundo

Analizando uno de los comentarios del periódico El Mundo (ver Ilustración 23), se encuentran dos palabras que transmiten alegría y una tristeza, con lo que predominaría la alegría. El valor del sentimiento indica, sin embargo, que se trata de un comentario negativo. Al comprobar el contenido del texto, se puede observar que se trata de un comentario irónico, de ahí el fallo en el análisis de emociones. El análisis de la ironía en textos es un tema complejo que aún no ha sido resuelto satisfactoriamente, pese a los esfuerzos realizados en esa dirección [6]

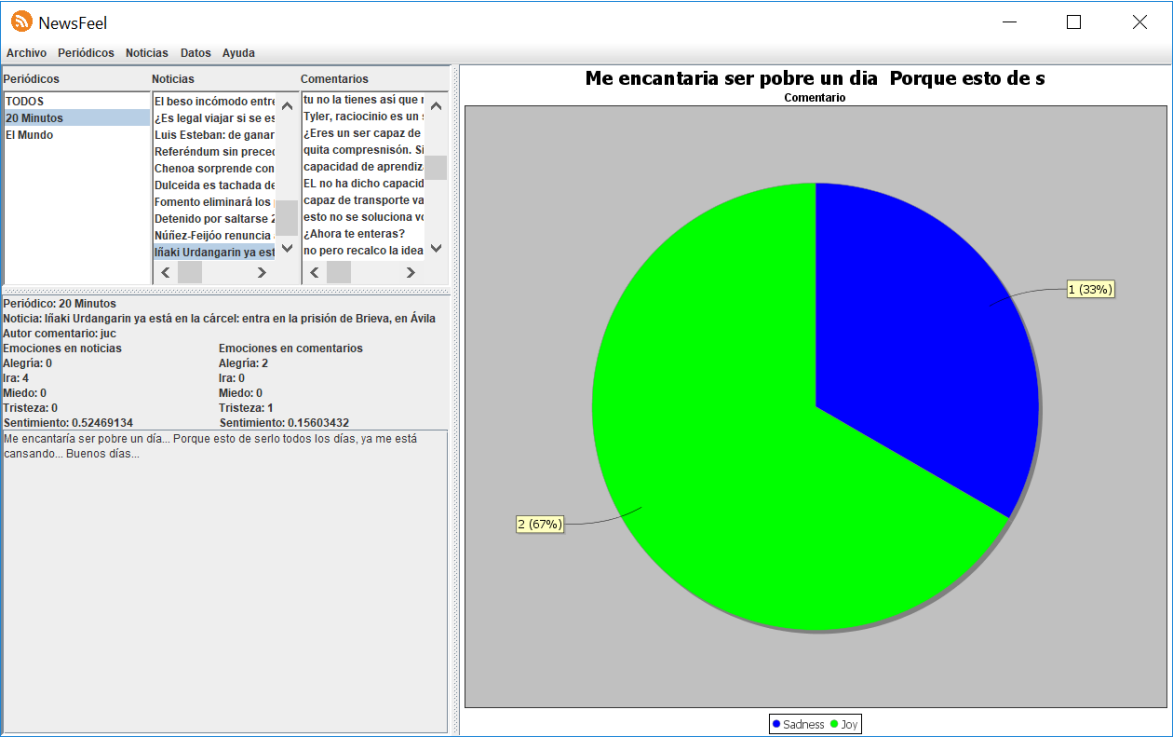


Ilustración 23. Datos de comentario del periódico El Mundo

7 Pruebas y resultados

Para comprobar la integridad de la aplicación se han realizado una serie de pruebas. Los casos probados, además de los casos nominales, engloban casos límites o casos de error que pudieran llevar a un error en la ejecución de la aplicación.

7.1 Pruebas funcionales

7.1.1 Pruebas de extracción de información

| Prueba | Comportamiento esperado | Resultado |
|---|---|-----------|
| Obtener una noticia utilizando una URL válida que pertenece a un periódico cualquiera del sistema. | Se obtiene la noticia. | OK |
| Obtener una noticia utilizando una URL de un periódico que no existe en el sistema. | No se obtienen noticias, se continúa la ejecución. | OK |
| Obtener varias noticias separadas por espacios y retornos. | Se obtienen las noticias correctamente. | OK |
| Obtener varias noticias no separadas por espacios. | No se obtienen noticias. | OK |
| Obtener una noticia con texto que no es una URL. | No se obtienen noticias. | OK |
| Obtener noticias por medio de RSS de un periódico seleccionado que contiene la funcionalidad de RSS. | Se obtienen todas las noticias disponibles de los RSS para ese periódico. | OK |
| Obtener noticias por medio de RSS de un periódico seleccionado que no contiene la funcionalidad de RSS. | No se obtienen noticias, se continúa la ejecución. | OK |
| Obtener noticias por medio de RSS de todos los periódicos. | Se obtienen noticias de los periódicos que tienen la funcionalidad RSS. | OK. |
| Obtener noticias por medio de búsqueda de un periódico seleccionado que no contiene la funcionalidad de búsqueda. | No se obtienen noticias. | OK |
| Obtener noticias por medio de búsqueda de un periódico seleccionado que contiene la funcionalidad de búsqueda. | Se obtienen noticias | OK |
| Obtener noticias por medio de búsqueda de un periódico seleccionado que contiene la funcionalidad de búsqueda, pero no existen noticias con esa palabras clave. | No se obtienen noticias | OK |
| Obtener noticias por medio de búsqueda en todos los periódicos habiendo algunos que no contienen la funcionalidad de búsqueda. | Se obtienen noticias. | OK |
| Obtener una noticia en la que el script no devuelva los datos esperados. | No se obtienen noticias. | OK |
| Obtener una noticia en la que el script tenga campos vacíos. | Se obtiene la noticia con los campos que estaban vacíos a un valor por defecto. | OK |

| | | |
|---|-----------------------------------|----|
| Obtener una noticia sin que exista el script a la hora de descargarla. | No se obtienen noticias. | OK |
| Obtener una noticia sin que el script consiga escribir un fichero con los datos. | No se obtienen noticias | OK |
| Obtener una noticia en la que el script devuelve un fichero con más datos de los esperados. | No afecta. Se obtiene la noticia. | OK |

7.1.2 Pruebas de interfaz

| Prueba | Comportamiento esperado | Resultado |
|--|--|-----------|
| Realizar pulsaciones en el menú, navegando entre todos los paneles. | Todos los paneles y acciones se muestran de manera correcta. | OK |
| Realizar selecciones en el panel de visualización, navegando por noticias y comentarios de todos los periódicos. | Se muestra la información correcta en todo momento. | OK |

7.1.3 Pruebas de gestión de periódicos

| Prueba | Comportamiento esperado | Resultado |
|--|---|-----------|
| Añadir un periódico de nombre único, con funcionalidades de RSS y búsqueda y ficheros de scripts existentes. | El periódico es añadido al sistema. | OK |
| Añadir un periódico de nombre único, sin funcionalidad de RSS ni de búsqueda. | El periódico es añadido al sistema. | OK |
| Añadir un periódico sin nombre. | Se muestra un error indicando que falta el nombre. | OK |
| Añadir un periódico sin URL base. | Se muestra un error indicando que falta la URL. | OK |
| Añadir un periódico sin script de descarga. | Se muestra un error indicando que falta el script de descarga. | OK |
| Añadir un periódico con un nombre de script de descarga inexistente. | Se muestra un error indicando que el script de descarga no existe. | OK |
| Añadir un periódico de nombre igual a otro que se encuentra en el sistema, con funcionalidades de RSS y búsqueda y ficheros de scripts existentes. | Se muestra un error indicando que ya existe un periódico con ese nombre. | OK |
| Añadir un periódico de nombre único, con funcionalidades de RSS y búsqueda y fichero de script de RSS no existente. | Se muestra un error indicando que no existe el fichero de script de RSS. | OK |
| Añadir un periódico de nombre único, con funcionalidades de RSS y búsqueda y fichero de script de búsqueda no existente. | Se muestra un error indicando que no existe el fichero de script de búsqueda. | OK |
| Añadir un periódico de nombre único, con funcionalidades de RSS y búsqueda y ficheros de scripts existentes, faltando la URL de RSS. | Se muestra un error indicando que falta el campo URL RSS. | OK |

| | | |
|--|---|----|
| Añadir un periódico de nombre único, con funcionalidades de RSS y búsqueda y ficheros de scripts existentes, faltando el script de RSS. | Se muestra un error indicando que falta el campo script RSS. | OK |
| Añadir un periódico de nombre único, con funcionalidades de RSS y búsqueda y ficheros de scripts existentes, faltando la URL de búsqueda. | Se muestra un error indicando que falta el campo URL búsqueda. | OK |
| Añadir un periódico de nombre único, con funcionalidades de RSS y búsqueda y ficheros de scripts existentes, faltando el script de búsqueda. | Se muestra un error indicando que falta el campo script búsqueda. | OK |
| Modificar el nombre de un periódico a uno que no exista en el sistema. | El periódico es modificado en el sistema | OK |
| Modificar un periódico borrando el nombre. | Se muestra un error indicando que falta el nombre del periódico. | OK |
| Modificar un periódico la URL base. | Se muestra un error indicando que falta la URL del periódico. | OK |
| Modificar un periódico borrando el script de descarga. | Se muestra un error indicando que falta el script de descarga. | OK |
| Modificar un periódico borrando el nombre. | Se muestra un error indicando que falta el nombre del periódico. | OK |
| Modificar un periódico cambiando el script de descarga a uno inexistente. | Se muestra un error indicando que el script de descarga no existe. | OK |
| Modificar un periódico con funcionalidad de RSS, eliminando la URL de RSS | Se muestra un error indicando que falta la URL de RSS. | OK |
| Modificar un periódico con funcionalidad de RSS, eliminando el script de RSS | Se muestra un error indicando que falta el fichero de script de RSS. | OK |
| Modificar un periódico con funcionalidad de búsqueda, eliminando la URL de búsqueda | Se muestra un error indicando que falta la URL de búsqueda. | OK |
| Modificar un periódico con funcionalidad de búsqueda, eliminando el script de búsqueda | Se muestra un error indicando que falta el fichero de script de búsqueda. | OK |
| Modificar un periódico con funcionalidad de RSS, poniendo un nombre de script de RSS inexistente. | Se muestra un error indicando que no existe el fichero de script de RSS. | OK |
| Modificar un periódico con funcionalidad de búsqueda, poniendo un nombre de script de búsqueda inexistente. | Se muestra un error indicando que no existe el fichero de script de búsqueda. | OK |
| Modificar el nombre de un periódico a otro igual que se encuentra en el sistema. | Se muestra un error indicando que ya existe un periódico con ese nombre. | OK |
| Eliminar el primer periódico cuando hay varios. | Se elimina el periódico correctamente. | OK |

| | | |
|--|--|----|
| Eliminar el único periódico disponible. | Se elimina el periódico correctamente. | OK |
| Eliminar el último periódico cuando hay varios. | Se elimina el periódico correctamente. | OK |
| Intentar eliminar un periódico cuando no hay más que borrar. | El sistema desactiva los campos y botones para evitar realizar las acciones. | OK |

7.1.4 Pruebas de almacenamiento y datos

| Prueba | Comportamiento esperado | Resultado |
|---|---|-----------|
| Guardar la base de datos en un fichero. | La base de datos se guarda correctamente en el fichero especificado. | OK |
| Guardar la base de datos en un fichero en otro directorio. | La base de datos se guarda correctamente. | OK |
| Guardar la base de datos con el mismo nombre que un directorio en la ruta actual. | La interfaz lleva hacia dentro del directorio. No se elimina el directorio. | OK |

7.1.5 Pruebas de análisis de emociones

| Prueba | Comportamiento esperado | Resultado |
|---|--|-----------|
| Se analiza una noticia emocionalmente con palabras de todas las emociones. | Se detectan todas las palabras. | OK |
| Se analiza emocionalmente una noticia sin palabras de emociones detectadas. | El número de palabras detectadas es 0. | OK |
| Se analiza un texto con carga emocional. | Se obtiene el sentimiento correctamente. | OK |

7.2 Pruebas no funcionales

7.2.1 Pruebas de rendimiento y estabilidad

Durante todas las pruebas anteriores no se ha observado ningún error o congelamiento de la aplicación. Los casos de error por los que se han pasado las pruebas anteriores no han hecho que la herramienta dejara de funcionar: en los casos pertinentes se mostró un error informando al usuario.

7.2.2 Pruebas de usabilidad

Durante las anteriores pruebas se han probado todos los elementos de la interfaz al completo, ejercitando los casos límite. No se han encontrado errores en interfaces, ni se han mostrado datos erróneos. Tampoco se han podido pulsar los componentes que en ese momento no tuvieran utilidad, habiendo sido desactivados.

Se han realizado pruebas de manera informal con familiares para comprobar la usabilidad de la herramienta. Se les pidió realizar una serie de acciones y posteriormente se les preguntó si habían tenido algún problema. Las acciones las realizaron correctamente y todos ellos comentaron que la herramienta fue sencilla de utilizar.

7.2.3 Pruebas de mantenibilidad

pensado en la separación de funcionalidad en todo caso. Prueba de ello es la flexibilidad del modelo de datos, pudiendo incorporar cualquier tipo de periódico, no sólo dentro de la herramienta sino con el uso de herramientas externas. Se ha probado incorporando, una vez que ya estaba funcionando con El Mundo, el periódico 20MinutosGenérico. Además, sería posible de manera sencilla incorporar también otros tipos de medios online (véase algunas sugerencias en el apartado de trabajo futuro), realizando pequeñas modificaciones en el modelo de datos, vista y controlador. También ha sido posible la incorporación de un nuevo indicador de sentimiento posteriormente a la utilización del análisis de emociones de textos sin grandes problemas.

8 Conclusiones

8.1 Conclusiones

Tras la finalización del proyecto se puede afirmar que se ha diseñado y desarrollado con éxito una herramienta genérica capaz de: incorporar de manera dinámica métodos de obtención de noticias y comentarios de periódicos online; analizar las emociones y el sentimiento que transmiten noticias y comentarios; y dar soporte a la visualización de toda la información, tanto la analizada como los resultados del análisis.

A continuación, se enumera el cumplimiento de objetivos especificados en la sección 1.2 del presente documento:

- Se ha conseguido facilitar la incorporación, modificación y eliminación de métodos de extracción de noticias.
- Se ha logrado obtener la información de las noticias de diferentes periódicos online
- Se obtiene la información de los comentarios de las noticias.
- Se ha logrado analizar emocional y sentimentalmente los textos de las noticias y comentarios.
- Se ha conseguido incorporar una interfaz para visualizar los datos y la información del análisis de emociones de las noticias y comentarios almacenados.
- Se permite comparar la información de los textos con el análisis de emociones de los mismos.
- Se permite guardar las representaciones de los datos de análisis de emociones.
- Se ha conseguido almacenar la información de las noticias y comentarios, así como los resultados del análisis de emociones.
- Se ha conseguido recuperar y volver a visualizar información extraída y analizada previamente.

Se han observado algunos aspectos interesantes tras el uso de la herramienta para analizar distintas noticias en distintos periódicos:

- Se han encontrado textos en los comentarios de las noticias en los que se han detectado unas emociones de felicidad, en principio positivas, mientras que el valor del sentimiento indicaba un texto negativo (ver Ilustración 23). En la mayoría de las ocasiones estos textos se trataban de figuras irónicas que el análisis de las emociones implementado actualmente no puede detectar.

La herramienta desarrollada se puede utilizar para analizar el *sentir* de la sociedad con respecto a temas de actualidad publicados en los periódicos online.

8.2 Trabajo futuro

Tras la realización de este proyecto han surgido algunas ideas sobre cómo se podría para mejorar el sistema dependiendo el enfoque que se considere necesario en el futuro.

El modelo de datos podría adaptarse para obtener no solamente periódicos, sino cualquier medio online. Actualmente se pueden añadir periódicos nuevos utilizando scripts específicos para descargar y obtener la información de las noticias. Se podría introducir una clase más genérica que abarcara todos los tipos de medios online y hacer aún más flexible la obtención de información. Esto permitiría añadir entradas y comentarios de blogs, hilos y comentarios de foros. En el caso de los blogs comparten mucho la estructura con los periódicos, al tener cada entrada un texto principal y comentarios. Sin embargo, en los hilos de los foros no hay un texto principal como tal, sólo comentarios. Podría entenderse que el primer comentario fuera en realidad el texto principal pero estructuralmente no es así. Además, los hilos habitualmente se estructuran en categorías, las cuales sería interesante analizar en conjunto.

Se podrían realizar búsquedas de noticias utilizando filtros. Para poder encontrar con la mayor facilidad entre todas las fuentes de información disponibles, una opción sería realizar búsquedas personalizadas, pudiendo indicar el número máximo de noticias que encontrar, obtener noticias que estén entre dos fechas, obtener sólo noticias que tengan en el texto de la noticia o en los comentarios una emoción específica mayoritaria o que los textos analizados tengan un sentimiento positivo o negativo. De esta manera la búsqueda sería más eficaz y tomaría más valor, al eliminar los resultados que no fueran relevantes para el usuario en ese momento.

Siguiendo en la misma línea, también se podrían implementar filtros en la representación de datos, para mostrar, de entre todas las noticias, las que cumplan una serie de reglas, como las utilizadas de ejemplo en el punto anterior.

Se podrían guardar palabras clave de cada noticia para elaborar una red de palabras clave y categorizar las noticias automáticamente, pudiendo utilizar las categorías como filtros o en las búsquedas.

Con respecto al análisis de emociones, se proponen las siguientes mejoras futuras:

- Analizar textos por oraciones en vez del texto completo. De esta manera no se condiciona cada parte del texto al análisis general de todo ello. Cada oración puede tener un sentido diferente en función de la intención del escritor y del contexto. También se puede estar hablando de dos temas diferentes en el mismo comentario.
- Identificación de locuciones o frases hechas que dan una intencionalidad distinta a los que significan las palabras que las contienen por separado. Dependiendo de cada una de ellas, al ser identificadas estas estructuras se deberá incorporar directamente una emoción o cambiar la emoción detectada.
- Ampliar los diccionarios existentes con sinónimos y antónimos. Se tendría que realizar un análisis para comprobar las tasas de acierto para validar las nuevas palabras. En el caso de los antónimos, pueden ser que las palabras con origen en un diccionario no tengan un diccionario de destino fijo y puedan ir en uno u otro dependiendo del caso.

Por otra parte, sería interesante detectar usuarios *tóxicos* que publiquen de manera regular en el periódico, posiblemente mensajes que nada tienen que ver con las noticias, o, aun estando relacionados, siempre en la misma dirección (p.ej., negativa).

- Desde el punto de vista de la interfaz, se podría formatear el texto analizado, marcando con colores las emociones detectadas para facilitar el estudio de la eficacia

del análisis de emociones llevado a cabo. En esta línea, se podrían exportar los resultados del análisis de emociones anotados, para permitir un análisis posterior y fuera del ámbito de la herramienta.

8.3 Consideraciones finales

La realización de este trabajo me ha supuesto una satisfacción tanto personal como académica, permitiéndome seguir mejorando en el ámbito elegido, aplicar los conocimientos adquiridos en el máster y superar el reto personal, técnico y de ingeniería que supone idear, planificar y realizar un proyecto de estas características.

Referencias

- [1] Ceron, A. (2015). Internet, news, and political trust: The difference between social media and online media outlets. *Journal of Computer-Mediated Communication*, 20(5), 487-503.
- [2] Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *Icwsm*, 7(21), 219-222.
- [3] Kasper, W., & Vela, M. (2011, October). Sentiment analysis for hotel reviews. In *Computational linguistics-applications conference* (Vol. 231527, pp. 45-52).
- [4] Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.
- [5] Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31, 527-541.
- [6] Barbieri, F., & Saggion, H. (2014). Automatic Detection of Irony and Humour in Twitter. In *ICCC* (pp. 155-162).
- [7] de Andrés Nogales, J. (2014). *Programa de soporte a la extracción de noticias y comentarios de lectores de periódicos online, y análisis de las emociones que transmiten* (Bachelor's thesis). Universidad Autónoma de Madrid.
- [8] ML Blog Team. (2015, April 8). Introducing Text Analytics in the Azure ML Marketplace. Recuperado de <https://blogs.technet.microsoft.com/machinelearning/2015/04/08/introducing-text-analytics-in-the-azure-ml-marketplace/>

A Manual de usuario

A.1 Instalación de programas asociados necesarios

A.1.1 PhantomJS y CasperJS

A.1.1.1 Instalación de PhantomJS (en Windows)

1. Descargar la última versión desde <http://phantomjs.org/>
2. Extraer el fichero «phantomjs.exe» en una ubicación conocida.
3. Añadir la ruta de la ubicación al PATH para que pueda ser ejecutado desde cualquier ubicación.

A.1.1.2 Instalación de CasperJS (en Windows)

1. Descargar la última versión desde <http://casperjs.org/>
2. Extraer el fichero «casperjs.exe» en una ubicación conocida.
3. Añadir la ruta de la ubicación al PATH para que pueda ser ejecutado desde cualquier ubicación.

A.2 Utilización de la herramienta

Para realizar las diferentes acciones que dispone la herramienta se ha de utilizar un menú, situado en la parte superior de la misma. Las acciones disponibles son las descritas a continuación.

A.2.1 Sistema

A.2.1.3 Base de datos

La base de datos está almacenada en un fichero llamado “noticias.xml” en el mismo directorio donde se ejecuta la herramienta.

A.2.1.4 Crear una nueva base de datos

Si no existe una base de datos previa se creará una automáticamente. En caso contrario, será necesario eliminar la existente borrando el fichero «noticias.xml» antes de ejecutar la herramienta. Tras ello, al abrir la herramienta la nueva base de datos será creada.

A.2.1.5 Guardar una base de datos

Se puede guardar el estado actual de la base de datos. Para ello pulsar en el menú «Archivo/Guardar BBDD». Al hacerlo se mostrará una ventana en la que se podrá dar un nombre al fichero y al pulsar el botón «Guardar» se escribirá el fichero.

A.2.1.6 Cargar una base de datos

Para cargar una base de datos previamente creada se deberán realizar los siguientes pasos:

1. Cerrar la aplicación (si fuese necesario).

2. Modificar el nombre o eliminar la base de datos actual.
3. Cambiar el nombre de la nueva base de datos a «noticias.xml»
4. Abrir de nuevo la aplicación.

Tras estos pasos los datos estarán en la aplicación de nuevo.

A.2.2 Periódicos

Además de los que se encuentran en la herramienta por defecto, 20 Minutos y El Mundo, se pueden añadir otros periódicos.

A.2.2.7 Añadir un periódico

Para entrar en la pantalla de añadir un periódico se ha de pulsar en el menú «Periódicos/Añadir periódico». Se mostrará la pantalla que aparece en la Ilustración 24.

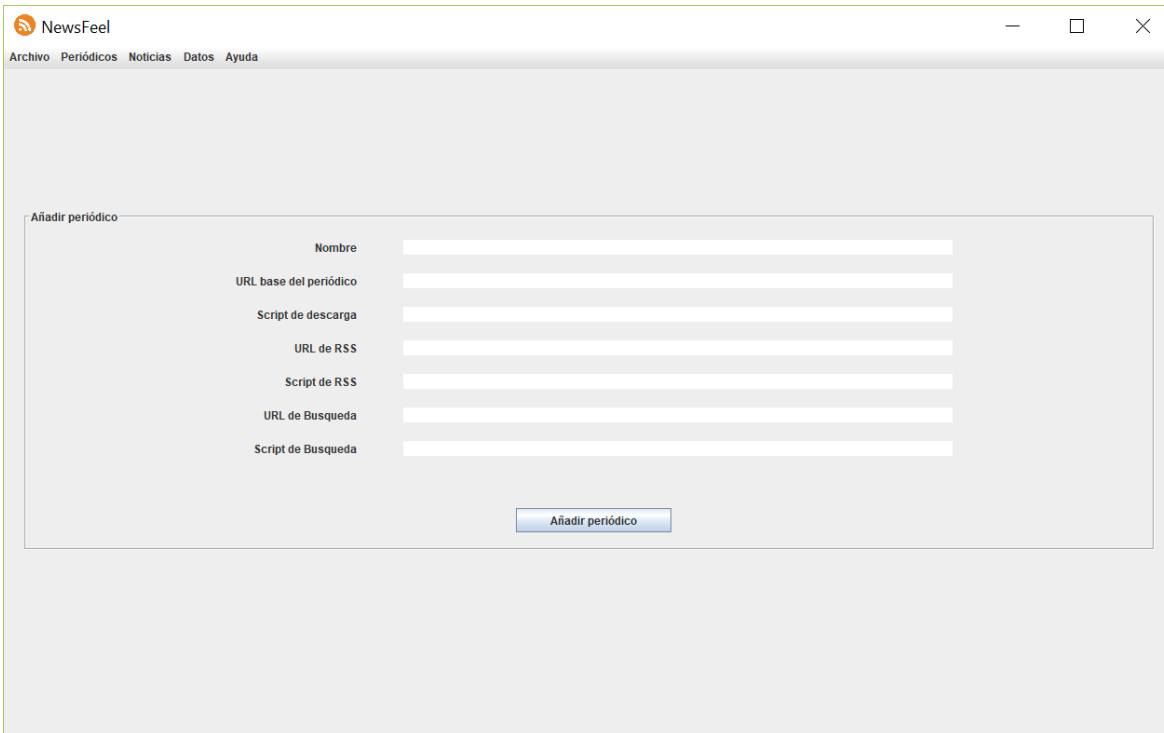


Ilustración 24. Añadir un periódico

Se han de rellenar los campos «Nombre», «URL base del periódico» y «Script de descarga» obligatoriamente para todos los periódicos. Como restricciones, el nombre no puede coincidir con otro periódico existente en la herramienta.

La URL base del periódico es lo que utiliza la herramienta para identificar que una dirección URL pertenece a un periódico u otro. Por esta razón se recomienda que sea la parte izquierda común de las URL de noticias del periódico. En la Ilustración 25 se muestra un ejemplo para el periódico 20 Minutos.

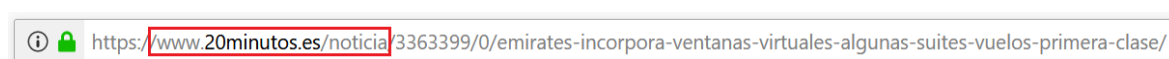


Ilustración 25. Ejemplo de URL base para el periódico 20 Minutos

El fichero de script de descarga debe existir en el mismo directorio donde se ejecuta la herramienta. Este fichero debe ser un script escrito para CasperJS que tras ser ejecutado genere un fichero con los datos de la noticia, como se describe en la sección 4.5.3.2 Script para periódicos genéricos.

Opcionalmente y si el periódico elegido dispone de ello se pueden rellenar los campos relativos a los RSS. Si se utiliza esta opción se han de rellenar necesariamente los campos «URL de RSS» y «Script de RSS». La URL de RSS se utiliza para obtener las últimas noticias que publica el periódico. El fichero de RSS debe ser un script para CasperJS que tras ser ejecutado genere un fichero con los enlaces de noticias obtenidos, como se describe en la sección 4.5.3.2 Script para periódicos genéricos. El fichero de script de RSS debe existir en el mismo directorio donde se ejecuta la herramienta.

Si está disponible en el periódico elegido también se puede añadir, también de manera opcional, los campos relativos a la búsqueda de noticias. Si se utiliza esta opción se han de rellenar necesariamente los campos «URL de Búsqueda» y «Script de Búsqueda». La URL de búsqueda se utiliza para obtener las noticias más relevantes con respecto a unas palabras clave usadas como búsqueda en la herramienta. Debe contener la URL hasta el símbolo ‘=’ que aparece cuando se realiza una búsqueda en un periódico.

Por ejemplo, al buscar noticias sobre la Comunidad de Madrid en el periódico El Mundo aparece lo que se muestra en la Ilustración 26. La URL de búsqueda para este periódico sería la marcada en rojo.

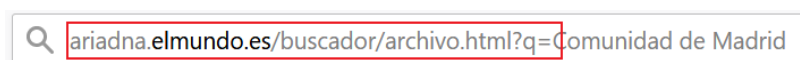


Ilustración 26. URL de búsqueda

El fichero de script de búsqueda debe estar escrito para CasperJS y, tras ser ejecutado, ha de generar un fichero con los enlaces de noticias obtenidos, como está descrito en la sección 4.5.3.2 Script para periódicos genéricos. El fichero de script de búsqueda debe existir en el mismo directorio donde se ejecuta la herramienta.

Para finalizar, se ha de pulsar el botón “Añadir periódico”. Un mensaje informará si el proceso ha tenido éxito o, en caso de error, si es necesaria alguna acción adicional.

A.2.2.8 Modificar un periódico

Para entrar en la pantalla de añadir un periódico se ha de pulsar en el menú Periódicos/Modificar periódico.

The screenshot shows a web application window titled 'NewsFeel'. The menu bar includes 'Archivo', 'Periódicos', 'Noticias', 'Datos', and 'Ayuda'. The main content area is titled 'Modificar periódico'. Inside this area, there is a section 'Seleccionar periódico' with a dropdown menu currently showing 'El Mundo'. Below this, there are several labeled input fields:

| Label | Value |
|------------------------|---|
| Nombre | El Mundo |
| URL base del periódico | www.elmundo.es |
| Script de descarga | ElMundo.js |
| URL de RSS | http://estaticos.elmundo.es/elmundo/rss/portada.xml |
| Script de RSS | ElMundo_rss.js |
| URL de Búsqueda | http://ariadna.elmundo.es/buscador/archivo.html?q= |
| Script de Búsqueda | ElMundo_búsqueda.js |

At the bottom of the form is a button labeled 'Modificar periódico'.

Ilustración 27. Modificar un periódico

En esta pantalla se pueden modificar todos los campos del periódico. Para ello, primero se ha de seleccionar el periódico a modificar. Si no hay periódicos que modificar los campos aparecerán desactivados.

Al igual que en el apartado A.2.2.7 Añadir un periódico, se aplican las mismas restricciones en los campos de entrada.

Por último, pulsar el botón «Modificar periódico». Un mensaje informará si el proceso ha tenido éxito o, en caso de error, si es necesaria alguna acción adicional.

A.2.2.9 Borrar un periódico

Para entrar en la pantalla de borrar un periódico se ha de pulsar en el menú Periódicos/Borrar periódico.

En esta pantalla se pueden borrar todos los datos del periódico. Esto incluye todas las noticias, comentarios y emociones almacenadas.

Para borrar un periódico primero se ha de seleccionar el periódico a borrar. Acto seguido, pulsar el botón «Borrar periódico». Un mensaje de aviso se mostrará para informar al usuario de la acción. Finalmente, si se pulsa el botón de aceptar, el periódico y sus datos serán borrados.

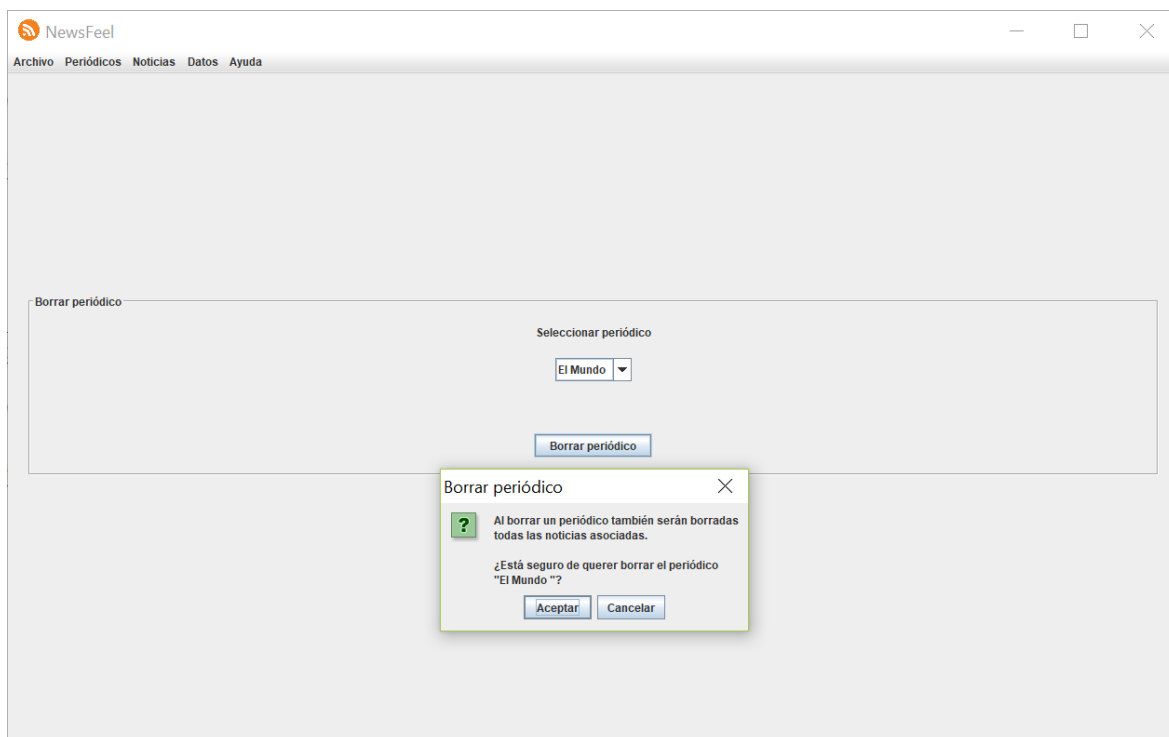


Ilustración 28. Borrar un periódico

A.2.2.10 Scripts de descargas

Los scripts de descargas sirven para obtener las noticias de los periódicos online. Al tener cada periódico una estructura distinta se han de realizar de manera específica para cada periódico.

Para obtener noticias desde distintas fuentes como URL, RSS o búsquedas se ha de seguir el formato de script en el apartado 4.5.3.2 Script para periódicos genéricos

A.2.3 Noticias

A.2.3.11 Añadir noticias

Para entrar en la pantalla de añadir noticias se ha de pulsar en el menú Noticias/Añadir noticias. Se pueden añadir noticias de tres métodos diferentes; mediante la URL de la noticia, a través de los RSS del periódico y mediante una búsqueda de palabras clave.

The screenshot shows the NewsFeel application window. It has a menu bar with 'Archivo', 'Periódicos', 'Noticias', 'Datos', and 'Ayuda'. The main content area is divided into three sections:

- Añadir noticias a partir de URLs:** A large text input field for entering URLs, with a button 'Obtener noticias desde URL' to the right.
- Añadir noticias a partir de RSS:** A dropdown menu labeled 'Seleccionar periódico:' with 'TODOS' selected, and a button 'Obtener noticias desde RSS' to the right.
- Añadir noticias a partir de palabras clave:** A text input field for keywords, a dropdown menu labeled 'TODOS', and a button 'Buscar noticias' to the right.

Ilustración 29. Añadir noticias

A.2.3.11.1 Añadir noticias a partir de URL

Para añadir noticias a partir de la URL de la noticia se deberán introducir las URL de las noticias separadas por espacios o retornos en el área de texto. Se pueden introducir URL de distintos periódicos a la vez; la herramienta se encargará de descargar los datos del periódico correspondiente. Para obtener las noticias pulsar el botón «Obtener noticias desde URL». A continuación, un mensaje aparecerá informando del número de noticias obtenidas y las noticias estarán disponibles en la herramienta para su visualización.

A.2.3.11.2 Añadir noticias a partir de RSS

Para obtener noticias a través de este método, seleccionar el periódico desde el cual descargar las noticias y pulsar el botón «Obtener noticias desde RSS». Es posible obtener las noticias de los RSS de todos los periódicos seleccionando la opción «TODOS» del desplegable. A continuación, un mensaje aparecerá informando del número de noticias obtenidas. Si un periódico no tiene la opción de RSS disponible no se obtendrán noticias de este periódico.

A.2.3.11.3 Añadir noticias a partir de búsqueda de palabras clave

Para obtener noticias a través de este método introducir las palabras clave en el área designada, seleccionar el periódico desde el cual descargar las noticias y pulsar el botón «Buscar noticias». También es posible buscar noticias de entre todos los periódicos seleccionando la opción «TODOS» del desplegable. A continuación, un mensaje aparecerá informando del número de noticias obtenidas. Si un periódico no tiene la opción de búsqueda disponible no se obtendrán noticias de este periódico.

A.2.4 Datos

A.2.4.12 Visualizar datos

Para entrar en la pantalla de análisis de datos se ha de pulsar en el menú Noticias/Analizar datos.

En esta pantalla se puede ver toda la información almacenada en la base de datos relativa a los periódicos, las noticias y comentarios que contiene cada uno, así como la representación de las emociones que los textos de las noticias y comentarios contienen.

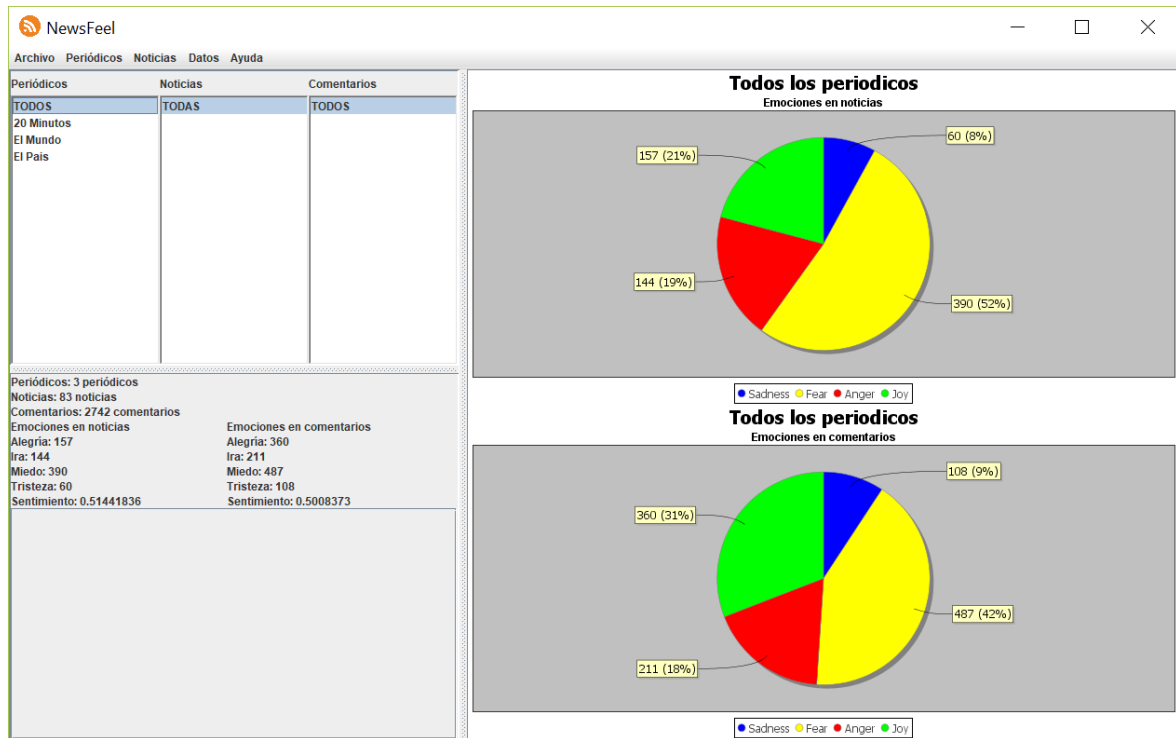


Ilustración 30. Panel de visualización

En función de la selección realizada en el panel de arriba a la izquierda se mostrarán los datos en el panel de la abajo a la izquierda. También se mostrarán distintos tipos de gráficos en el panel derecho en función de esta selección.

Los diferentes tipos de selección son los siguientes:

- Todos los periódicos.
- Un periódico y todas sus noticias.
- Un periódico, una noticia y todos sus comentarios.
- Un periódico, una noticia y un comentario.

A.2.4.13 Guardar imágenes de resultados

Es posible obtener las imágenes de los resultados y guardarlas para un uso posterior. Para ello basta con pulsar con el botón derecho en cualquier imagen que aparezca en la pantalla de visualización de datos, seleccionar la opción «Save as» y pulsar en cualquier de las tres opciones dadas; PNG, SVG o PDF, como se muestra en la Ilustración 31.

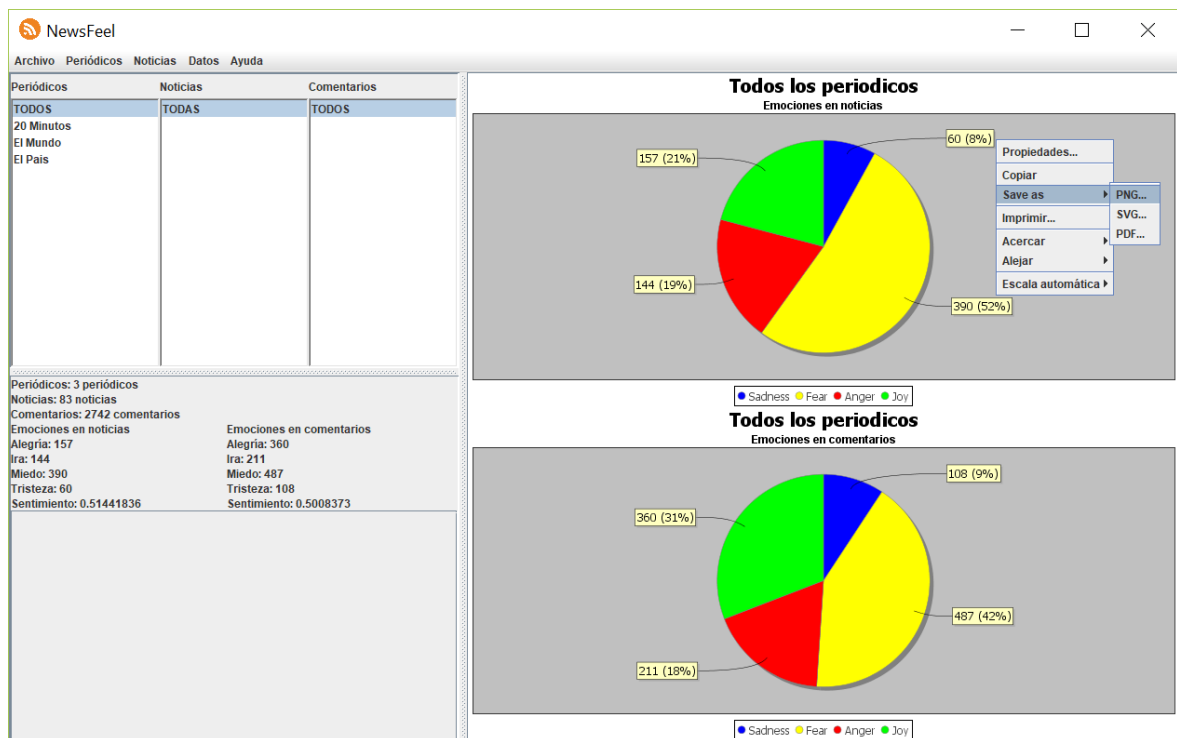


Ilustración 31. Guardado de imágenes de resultados

